PRICEWATERHOUSECOOPERS

RETAIL SALES TAX COMPLIANCE COSTS: A NATIONAL ESTIMATE

Volume Two: Survey Documentation

Prepared for

Joint Cost of Collection Study

Prepared by

National Opinion Research Center
University of Chicago

April 7, 2006

RETAIL SALES TAX COMPLIANCE COSTS: A NATIONAL ESTIMATE

Table of Contents

Volume One: Main Report

Volume Two: Survey Documentation

I. SURVEY METHODOLOGY

A. Introduction

This is the summary report of the NORC documentation provided to complete the work done for PricewaterhouseCoopers (PwC) on the 2005 State Sales Tax Compliance Cost Study.[1]

The material is organized into 8 subsections. A narrative approach is taken beginning with the sample design and implementation for the Study (Subsection I.B). Questionnaire development is covered briefly in Subsection I.C. Field implementation issues are covered in Subsection I.D, especially efforts to improve the response rate. Cleanup work to handle question misunderstandings and data gaps comes next (Subsections I.E and I.F). The way survey estimates were made completes our narrative (Subsection I.G). In a concluding section (Subsection I.H), we provide a statement of the limitations we see in the work done so far, and recommend techniques that PwC might wish to employ in their later analytic use.

Supporting this summary are three more documents: The codebook for the survey database produced is given in Section II, data checking details including counts of the inconsistencies found are shown in Section III, and the standard error analysis is presented in Section IV.

The focus of all of these documentations is not simply to describe what was done -- albeit that is clearly our first goal. Rather we look at how well the Study resources in time, money and opportunity costs were used to handle the four major challenges or errors that confront all survey work: coverage error, missing data, measurement error and sampling error.[2] Of the four only the impact of sampling errors can be quantified fully. Coverage errors depend on the frame or list that is available for use or that can be constructed. Measurement errors depend on the instrument or questionnaire used, the willingness of respondents to consult records, etc. Finally missing data depends both on the interview nonresponse rate and, secondarily, on the instances where, even though a questionnaire was returned, some items were left unanswered and have had to be imputed.

[1] Also called the Joint Cost of Collection Study (JCCS).
[2] For example, see Deming, W.E. (2006), ***The American Statistician***.

It is our view that, if employed properly, this first-ever National Study of state sales tax compliance costs is “fit for the uses” for which it was intended. This is so, despite important limitations that we treat in detail below. This point will be returned to in the concluding remarks of this Section.

B. Sample Design and Implementation

B.1 General. The population for the Study sample consists of retail firms (excluding "eating and drinking establishments") with known sales above $150,000. All companies on the Duns (or D&B) file with sales greater or equal to $100,000,000 were selected with certainty. The remaining companies were sampled proportionally within eight and then finally nine state groupings with a minimum of 500 companies sampled per state bucket or grouping (Table 1 shows the composition of each of the final nine state groupings).

Table 1. Final States within Each of the Nine State Groupings or Buckets

State Group 1	State Group 2	State Group 3	State Group 4	State Group 5	State Group 6	State Group 7	State Group 8	State Group 9
CT DC HI IN KY MA MD ME MI MS NJ RI WV	PA VA VT	IA MN NC WI	AR IL MO NM OH TN TX UT WA	FL KS NE NV OK WY	CA GA SD	ND NY SC	AK AL AZ CO ID LA	DE MT NH OR

Table 2. Population and Sample by the Eight Original State Groupings or Buckets

State Grouping	Population Size (Rounded)	Total Sample Size	Certainty Stratum	Sample Left to Allocate
Group 1: No Local Sales Tax	255,000	2,000	240	1,760
Group 2: Uniform base and rate, origin-based	100,000	800	107	693
Group 3: Uniform base and rate, destination-based	103,000	800	115	685
Group 4: Uniform base, variable rate, origin-based	321,000	2,500	342	2,158
Group 5: Uniform base, variable rate, destination-based	485,000	3,400	456	2,944
Group 6: Variable base, variable rate, state administration	64,000	500	41	459
Group 7: Variable base, variable rate, local administration	41,000	500	40	460
Group 8: No Sales Tax	41,000	500	37	463
Total	**1,410,000**	**11,000**	**1,378**	**9,622**

Note: The numbers here are an approximation, simply using the totals from the D&B tables. This includes the unknown sales and under $150,000 in sales companies – the assumption, which at least holds true for unknown sales, is that the percentages of companies outside the sampling population is constant across the state groupings. There were 644,602 firms of $150,000 or more in retail sales, which serve as our final Study population.

Table 3. Original Eight State Grouping Study Sample Design

Size Class ($M)	State Grouping								Overall Total
	1	2	3	4	5	6	7	8	
.15 – .2	104	40	41	174	214	33	30	31	667
.2 – .5	363	135	130	435	662	105	94	95	2,019
.5 – 1.0	118	46	45	136	192	32	31	34	634
1.0 – 2.5	181	68	70	195	286	44	46	52	942
2.5 – 5.0	128	50	55	151	214	34	37	38	707
5.0 – 10.0	151	63	65	167	224	38	34	43	785
10.0 – 25.0	270	121	115	318	402	59	62	69	1,416
25.0 – 50.0	241	98	91	304	392	62	59	59	1,306
50.0 – 100.0	204	72	73	278	358	52	67	42	1,146
100.0 + (Certainty)	240	107	115	342	456	41	40	37	1,378
Total	**2,000**	**800**	**800**	**2,500**	**3,400**	**500**	**500**	**500**	**11,000**

Note: This is the stratification as the end of August, 2004. See Table 4 for the final groupings.

Table 2 shows the breakdown of sales by the original eight state groupings. The original 11,000 sample selections were further stratified by sales class within each group or bucket. Compliance costs are estimated for the sales classes, and the sample was allocated proportionally according to the estimate. Table 3 above shows the various sales classes and the originally allocated sample size. See Table 4 for the final sample sizes.

B.2 Duns File Coverage and Nonresponse Concerns. Data problems exist with the Duns file and nonresponse is expected to be sizeable. Some points that bear here are:

1. Whether to sample cases below the $150,000 threshold or to include cases with an unknown sales amount? In the end we did not include such cases in the selection. Given inflation, the real focus of the results probably should be firms with $200,000 in sales anyway.

2. How to handle dated sales and retail sector information? The dated nature of the Duns file was partly self-correcting, using the survey results. We also adjusted further by reweighting the sample cases to external Census Bureau totals for the retail sector.

3. What to do to assure that the expected large nonresponding fraction would not bias our results or damage the Study's credibility? We obviously could not insure against bias arising because of the nonresponse. We did partially measure it and did, as will be discussed below, adjust for it. Further, we acquired some tips, so as not, at least, to misuse the data that were obtained.

4. Econometrically, since the "no sales tax" states (group 8) will be used as a standard to compare against the others, should the sample size there be larger? Originally NORC recommended that the "no sales tax" states have their sample size doubled, as the analysis relies so heavily on this comparison. However, in the end, this was not done. Instead, we opted to raise the overall sample by an additional 2,000 selections, so as to simultaneously deal with the smaller that expected response while at the same time imbedding a bias study into the overall project.

5. Should additional cases be selected to guard against differential nonresponse across groups? Depending on the state-to-state group comparisons some sample size insurance, say a 10% rise, might be warranted for the groups 6 and 7. Of course, this would mean a drop elsewhere but that cost should not be important for most analyses. In the end, this step was not taken. See No. 4 above. In the same vein was a proposal to oversample direct marketers. This too was not done, as the direct marketers withdrew their support for the effort.

At the design stage, NORC remained worried about a smaller than desired sample size and sought all suggestions that could ameliorate this. Among other steps, a large pilot test, not called

for under the contract, was conducted to this end. The pilot tested a number of ideas, but mainly whether dividing the sample up by size into the smaller firms that would get a short questionnaire and larger firms that would get a more complete version. In the end only a single questionnaire was fielded (See Subsection I.C below).

B.3 Initial Expected Sample Yields. Previous NORC surveys using the D&B business lists had shown that multiple contacts and an instrument that is not too burdensome can yield response rates of around 30%, once inactive cases on the frame (between 10% to 25%) were excluded and a small incentive was paid.[3] Budgets did not permit an incentive, so we expected response perhaps to be halved, ranging from 15% to 20%. The math here means that the expected sample sizes, subject to considerations mentioned below, would range from just over 1200 to just under 2000. When the decision was made by PwC to use only a single longer instrument the likely response rate should have been halved again. Anyway that was about where we ended up, at roughly an 8% (raw) response rate.

In any case we were worried about the burden in the survey instrument. For the longer version, several hours appeared to be required for modest sized businesses and potentially much more for the very largest firms. One key technique, we did employ, was to send the questionnaire to the attention of the tax department. We are convinced that this did reduce handling time but many potential respondents who originally agreed to return the survey, did not. We cannot say with any certainly how many of these firms started the task and then gave up, versus firms that simply answered "Yes," never having any intention of complying (i.e., a soft "No").

B.4 Actual Sample Yields. During the Study period the classification into the original 8 state buckets was refined and a ninth category was added in September of 2004. This was the way the actual survey began in October of 2004, with a target of 11,000 cases to be contacted.

However, as our concerns about a low response rate were increasingly borne out, we added in November almost 3,000 more selections to see if we could increase the respondent sample size.

[3] The Small Business Survey work done for the Federal Reserve Board in 1998 and again in 2003 was the basis for this observation. It should be noted that response rates, for a given level of effort fell greatly between the two

These further selections came for an earlier survey and were businesses that were known to be still on the D&B frame.[4] See Table 4 below.

Table 4. Final Nine State Grouping Study Sample Size

Sales Class	State Group									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Total**
150k - 200k	129	49	57	39	90	146	77	57	93	737
200k - 500k	440	167	183	716	276	403	265	182	287	2,919
500k - 1M	148	56	62	191	81	124	73	59	104	898
1M – 2.5M	220	80	97	270	115	175	105	79	138	1,279
2.5M - 5M	146	59	66	182	71	117	78	56	102	877
5M - 10M	160	68	78	201	74	121	81	60	114	957
10M - 25M	286	127	138	372	133	224	140	97	182	1,699
25M - 50M	254	100	109	348	140	234	112	98	135	1,530
50M - 100M	211	69	85	310	126	221	87	97	47	1,253
100M+	312	121	144	462	169	237	136	104	38	1,723
Total	**2,306**	**896**	**1,019**	**3,091**	**1,275**	**2,002**	**1,154**	**889**	**1,240**	**13,872**

While these last selections were sent out too late to go through all the follow up steps described in Subsection I.D, they did add some more cases and gave us a way to look at potential nonresponse bias. More will be said about this next when we discuss the raw response rate achieved and attempt to give it an interpretation (See Subsection I.G).

What, then, was the overall response achieved? After all the steps taken there were about 800 questionnaires returned (796 in fact). Table 5 displays the response by state bucket and D&B sales class.

surveys and even though the 2003 survey had a comparable final response rate, the field cost per case rose greatly, perhaps doubling.

[4] The reason for this is that it simply took too long to negotiate the needed arrangement to fully integrate these cases into the main study. Even so because these businesses had been surveyed earlier it was possible to use them to separate the nonresponse rate into an ignorable and a potentially nonignorable portion. This terminology is related to the degree of potential residual bias that may be present. See Scheuren (2005a, 2005b).

Table 5. Final Count of Survey Responding Firms

Sales Class	State Group									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Total**
Under 150k	1	-	1	1	1	-	-	-	-	**4**
150k – 200k	3	2	1	-	2	1	5	1	1	**16**
200k – 500k	10	10	7	25	7	9	8	9	12	**97**
500k - 1M	7	5	2	14	6	7	1	-	8	**50**
1M - 2.5M	18	5	10	21	7	8	6	3	14	**92**
2.5M - 5M	16	3	6	27	5	14	1	3	8	**83**
5M - 10M	11	1	2	18	6	9	7	4	9	**67**
10M - 25M	17	13	12	23	6	13	4	6	13	**107**
25M - 50M	18	2	9	26	5	14	10	6	5	**95**
50M – 100M	10	5	7	19	12	7	3	8	1	**72**
100M+	25	6	12	32	11	8	7	9	3	**113**
Total	**136**	**52**	**69**	**206**	**68**	**90**	**52**	**49**	**74**	**796**

Note that this Table shows the response per the D&B frame variables. However, in the end, when the respondent indicated sales figures higher than the frame variable those were used instead. While no companies under $150k in sales were sampled, one response that could not be matched to the frame indicated sales below $150k. Three companies who did not indicate sales nor could be matched to the frame are also included here.

C. Questionnaire Development

Questionnaire development began in July of 2004 with a review of existing instruments created in earlier state-level sales tax compliance cost studies, notably a 1998 study done by Washington State and covering both Washington and Oregon. The actual questions to be asked in this National Study went beyond the earlier work, though, and had an independent foundation in a thoroughly developed economic theory. The actual process of question construction was thus guided by both practice and theory.

Naturally, in such an exercise, there are compromises to be made. Not everything that is wanted for econometric modeling may be practical to collect. In our National Study we began with a questionnaire that was well over 8 pages and was believed by some of the researchers on the development team to require respondents to extensively consult records. The respondent burden

of such an approach, unless respondents were paid (not an option), was viewed as excessive and likely to lead to a very low response rate.

During early August of 2004 two versions of the questionnaire were fielded on a pilot basis to see what the answers would be like and what the differential response rate might be. There was a four-page version and an eight-page one. At the time we were toying with the idea of using the long version only for the largest retailers; the short version for the rest. It turned out that neither version did very well, although the shorter version had much a higher response rate.

Two actions were taken after the poor performance of the pilot became clear. Each of these is taken up briefly:

> First, there were major efforts to streamline the questionnaire, smoothing out individual questions, looking closely at question order, and eliminating extraneous, nice-to-have but nonessential questions. While the survey practitioners on the team remained concerned about instrument length, eventually a single version emerged, to be given to everyone. By now it was October of 2004 and, while a second pilot might have been desired, there was simply neither the time nor the money to conduct it.

> Second, in parallel with the questionnaire development, there were also important efforts to produce introductory and reminder letters that would motivate higher response, to obtain extensive endorsements from the largest retailer associations, and, even, to get a news article placed in a trade journal. All of these were quite successful, it might be noted. They still did not seem to have outweighed, however, the perceived burden that the final questionnaire imposed. Nonetheless, we view the work here as a clear plus.

Arguably the balance between respondent burden versus analytic completeness might have been struck at a different point. We must await the analyses themselves, however, to determine what further streamlining might have been possible without sacrificing key results.

C.1 Questionnaire. The final questionnaire is eight pages in length. As can be seen from it, the cover page provides a map of the United States and lists the sponsoring organizations that endorsed the Study. The last page is primarily for follow-up and contact information; thus there are really only six pages of data requested.

Even though this data questionnaire can be viewed as of only modest burden, many potential respondents still did not return the form. As noted the response rate was only about 8%. This is so despite an extensive effort during field implementation to secure cooperation. See Subsection I.D for more on this.

C.2 Questions. The Study questions were tested as part of the pilot and also revised after being examined by the Study JCCS Steering Committee and by using focus groups (See Exhibit 1 for the note used to secure focus group participants). Still, there is evidence that some respondents misunderstood what was wanted. For the details here, Section III of this Volume might be consulted.

We did not find, it might be stressed, any evidence that there were wholesale misunderstandings of the questions. To underline this point, just 34 of the 796 questionnaires received were so incomplete as to be unusable.

It is clear, however, from the nature of the checking done later that respondents, in many cases, tried to fill out the questionnaires quite quickly and that sloppiness and outright gaps occurred that had to be fixed up with callbacks and imputations. [5]

D. Field Survey Operations

Survey field operations began in earnest after the pilot phase ended in September of 2004. The basic approach was to mail out the survey in two broad waves, depending on the time zones that the selected cases were located. Eastern Time Zone cases were where we started, then Midwest, Mountain and Western.

In this initial step, as soon as the mail sent was expected to arrive, a prompting call would be made from the NORC Call Center in Chicago. The basic letter used is shown as Exhibit 2.

[5] Elsewhere (Subsection I.H), NORC has recommended that for key analyses one or more alternatives to the hot desk imputations done be tried, using a deeper degree of subject matter knowledge.

In any case, the Call Center was to work each time zone's mailing in turn and then do a second round and ultimately in part a third round of prompting to elicit response. This approach was employed to be sure that we got good coverage of the country and, at the same time, were able to add cases, to the extent resources permitted. There were many special steps taken after or alongside the first set of contacts. Two are mentioned below:

> When it became obvious that businesses that promised to send in their completed questionnaires were not doing so, a second round of calls was made to those businesses that had promised to respond. Faxed copies of a blank questionnaire were provided to those who requested them.
>
> For the 50 or so largest businesses, many of which where sponsoring the Study, we made even more of an effort. Multiple contacts were made, in fact. Still the response was very disappointing. So, with the help of the JCCS Steering Committee, NORC was able to locate a better contact inside each large firm and to provide that firm with yet another blank questionnaire – by mail this time.

Exhibit 1. Focus Group Letter Used to Obtain Survey Reviewers

NATIONAL RETAIL FEDERATION

MEMORANDUM

THE VOICE OF RETAIL WORLDWIDE

Liberty Place, 325 7th Street, NW, Suite 1100,
Washington, DC 20004
Phone: 202.783.7971 Fax: 202.737.2849

I am writing to invite you to participate in a public/private sector study designed to measure the cost of collecting state and local sales tax in the United States.

NRF is the world's largest retail trade association with membership that includes department, specialty, discount, catalog, Internet, and independent stores. NRF supports the work of the Streamlined Sales Tax Project, an effort to simplify and modernize sales and use tax collection and administration.

As part of the Streamlined Sales Tax Project, a public/private sector Joint Cost of Collection Study group with public and private representatives (including NRF) was formed to determine the costs retailers incur to collect state and local sales taxes (see, www.streamlinedsalestax.org). NRF believes retailers should be compensated for the cost of collecting sales taxes; however, no nationwide study of the cost of collecting these taxes exists. The purpose of the cost of collection study is to develop an authoritative and independent measure of collection costs as well as an analysis of the potential impacts of the Streamlined Sales Tax Project on these costs.

The accounting firm of PricewaterhouseCoopers LLP (PwC) has been retained by the Joint Cost of Collection Study to undertake the measurement of collection costs. For this project, PwC has teamed with the Office of Tax Policy Research at the University of Michigan and the National Opinion Research Corporation at the University of Chicago.

PwC is seeking retailers who would be willing to help the study team develop a questionnaire that will be mailed to retailers across the country as part of the study. If you would be willing to test the survey questionnaire and participate in a webcast discussion regarding the questionnaire, ***please return the enclosed, self-addressed stamped postcard.***

Your participation will help assure that the cost of collection accurately measures all significant costs incurred by retailers as part of their sales tax collection responsibilities. All information collected as part of this study will be held in strict confidence by PricewaterhouseCoopers. For more information, please contact XXXX directly by phone (YYYY) or email (ZZZZ).

Exhibit 2. Basic Survey Questionnaire Introductory Letter

Tax Department
[company]
[street]
[city, state zip]

Dear Sir or Madam:

The enclosed survey is being conducted to determine how much U.S. retailers should be compensated for their costs of collecting sales tax and how the burden of the sales tax can be reduced. The survey is jointly sponsored by the following business and government organizations:

- National Retail Federation
- Direct Marketing Association
- Council on State Taxation
- Multistate Tax Commission
- Federation of Tax Administrators
- National Conference of State Legislatures
- Government Finance Officers Association

In addition to the above sponsors, the enclosed retailer questionnaire has been endorsed by:

- U.S. Chamber of Commerce
- Retail Industry Leaders Association
- National Federation of Independent Business

The questionnaire will be used to develop the first national measure of sales tax compliance costs. It will be an objective source of information for policymakers considering the extent to which retailers should be compensated (through discounts or allowances) for their collection costs.

PricewaterhouseCoopers LLP, a national accounting firm, has been charged with administering the questionnaire. Your business was selected using a scientific sampling process that gave retailers in each category an equal chance of being selected. Only a few retailers received this questionnaire. Thus, it is important that as many retailers return it as possible.

The information you provide is **strictly confidential**. Only researchers at the National Opinion Research Center (NORC) at the University of Chicago will see your individual data. No government agency, business organization, or any other party will have access to your response.

Thank you for your prompt attention to this important effort. If possible, we would appreciate your assistance in returning the questionnaire within one seven (7) business days of receipt.

Sincerely,

Peter Merrill

Peter Merrill
Principal

Exhibit 3. Basic Script for NORC Prompting Calls

[IF THE NUMBER IS INVALID, RECORD AS INVALID]

[IF ANSWERING MACHINE]

Hello. I'm with the National Opinion Research Center at the University of Chicago calling on behalf of the National Retail Federation and the Multistate Tax Commission. We're calling regarding an important survey recently sent to the Tax Department of your business regarding the collection of sales tax. If you have the survey form, please take the time to complete it. Your participation will aid in a national effort to reduce the costs to retailers and states of collecting and administering state and local sales tax. If you did not receive the survey, it can be downloaded at *www . jccs . info* or you can call us toll-free at [PHONE CONTACT] to request a copy. That number again is [PHONE CONTACT]. Because your response is important, we will follow-up in a week or so if we don't hear from you. Thank you in advance for your participation.

[IF PERSON ANSWERS]

Hello, this is [INTERVIEWER] calling from the University of Chicago on behalf of the National Retail Federation and the Multistate Tax Commission. I'm trying to reach the Tax Department or the person most qualified to answer questions regarding state sales taxes.

[IF CONNECTED TO TAX DEPT/PERSON AND A MACHINE PICKS UP, LEAVE ABOVE MESSAGE]

[IF GATEKEEPER INDICATES SUCH A PERSON EXISTS BUT IS UNAVAILABLE]

I'm calling regarding an important National Study being conducted on the collection of state and local sales tax. Recently, we sent out the survey materials to your business and I wanted to make sure the materials were received. If [TAX PERSON] did receive the survey, we are hoping he/she can complete it and return it to us.

Would you please tell [TAX PERSON] that if he/she needs another copy of the survey form, it can be downloaded at *www . jccs . info* or he/she can call us toll-free at [PHONE CONTACT] to request a copy. Because his/her response is critical to this study, we will follow-up in a week or so if we don't hear from him/her.

[ADDRESS ANY QUESTIONS]

Thank you very much for your time. Good-bye.

[IF APPROPRIATE PERSON IS REACHED]

Hello, I'm calling from the National Opinion Research Center at the University of Chicago. Recently, we sent you a survey on the collection of state and local sales tax. I wanted to make sure you received this mailing. Do you recall getting such a survey in the mail?

[IF SURVEY WAS RECEIVED]

That's great. Would it be possible for you to complete and return it within the next week?

[IF THEY AGREE TO COMPLETE SURVEY]

Great! Because your response is critical to this study, we will follow-up with you in a couple weeks or so, in case we don't hear from you.

[IF THEY ARE RELUNCTANT TO COMPLETE SURVEY]

Your participation would aid in a national effort to reduce the costs to retailers and states of collecting and administering state and local sales tax; an effort that has the cooperation of over 40 state governments. I hope that you reconsider participating in this study. Thank you for your time. Good-bye.

[IF SURVEY WAS NOT RECEVED]

I'm sorry about that. I can send you the survey right away either by fax or mail, or you can download the survey off the internet. Which would you prefer?

[IF THEY AGREE TO COMPLETE THE SURVEY:
BY MAIL: GET THEIR ADDRESS
BY FAX: GET THEIR FAX NUMBER
BY INTERNET: GIVE THEM THE URL
www . jccs . info]

Thank you for your participation. Because your response is critical to the study, we will follow-up in a week or so, in case we don't hear from you. Thank you for your time. Good-bye.

[IF THEY ARE RELUNCTANT TO COMPLETE THE SURVEY]

Your participation would aid in a national effort to reduce the costs to retailers and states of collecting and administering state and local sales tax; an effort that has the cooperation of over 40 state governments. If you reconsider, you can download the survey at *www . jccs . info* – thank you for your time. Good-bye.

[RECORD THE RESULT OF THE CALL:

WAS THE NUMBER VALID?
IF YES, DID YOU GET A HOLD OF A PERSON?
IF YES, DID YOU GET IN TOUCH WITH A TAX INDIVIDUAL?
IF YES, DID THEY RECEIVE THE SURVEY?
DID THEY AGREE TO COMPLETE THE SURVEY?]

It would be nice to say that these extra steps led to a greatly improved response. While both did increase response, the changes were not dramatic. Still the response rate among the largest firms was the best in any strata, at 25%.

D.1 Telephone Disposition Results. Maybe one of the best ways to characterize what NORC's prompting calls were like is to see the basic script that callers were to use. This appears here as Exhibit 3. The call prompting of the largest firms (not shown) was more extensive and since frequent callbacks were made, as personalized as possible.

The way each prompting call was coded is shown in Table 6 below. Notice, first, that the total shown differs from the total in Table 4 by 454 companies – mostly due to the fact that not all sampled cases could be loaded into the prompting system because either a phone number was not available in the D&B database or the D&B number was a duplicate.

A second observation to make is that there were also cases prompted but not mailed, due to invalid or duplicate D&B addresses. We are using the prompting disposition codes here instead of the receipt (mailing) disposition codes because more information is available to us. Even with these restrictions, there were many sample businesses that could not be located or were not longer in operation.

D.2 Implications of Prompting Results on D&B Frame coverage. The results of the prompting calls allowed NORC to quantify the extent to which the D&B frame was out-of-date and incomplete. This was not a small problem either. The cost of this incompleteness is manageable, although expensive – just a smaller sample size.[6]

[6] Some of the codes are clearly indications that the business was no longer operating ("No Longer in Business" was coded 229 times) or that the information that was on the D&B frame was insufficient to locate the business ("Unlocatable" was coded 910 times). Others could also be an indication that the business was not operating ("Maximum number of calls reached" at 1897) would be one of these

Table 6. Telephone Prompting Disposition Codes

Disposition	Cases	In-Scope
No Longer in Business	229	N
Respondent Promises to send in Questionnaire	330	Y
Respondent Already sent in Questionnaire	288	Y
Prompting Successful -- no additional info on respondent's likelihood to respond	5,738	Y
Prompting Successful -- respondent reluctant	303	Y
Maximum number of calls reached	1,897	N
Line disconnected or always busy	33	N
Unlocatable	910	N
Unavailable -- Prompt unsuccessful	477	Y
Refusal from respondent	1,977	Y
Foreign Language – Spanish	3	N
Refusal from Gatekeeper	1,005	Y
Foreign Language -- Not Spanish	12	N
Hostile Refusal	9	Y
Other	207	Y
Total	**13,418**	***Mixed***

Not so simple to handle is that there were businesses unknown to D&B or firms created after the frame was last updated that had yet to be posted and, hence, were unavailable for potential selection – this leading to some undercoverage. These were a more serious concern and led PwC to obtain an independent total of in-scope retail trade businesses from the Census Bureau and the Bureau of Labor Statistics.

The D&B frame NORC used was supposed to identify enterprises and not just their establishments. This did not always have happen leading to large firms, having many chances of selection. We, of course, were on the lookout for cases where the amounts reported in the survey were much larger that those on the D&B frame and made an adjustment for them, as described in Subsection I.G.

E. Consistency of Data

Consistency checks, largely specified by PwC, were performed on the database, leading to several discovered seeming inconsistencies in the data. These are tabulated in Section III of this Volume. That section also presents marginal information, by consistency check, for both failure counts and missingness counts. Note that the missingness is not a count of the number of missing survey items – it is simply a count of consistency checks with too much missing information to perform an imputation.

Companies were contacted to determine causes for some of these inconsistencies, and PwC recommended certain actions for others. Where a company was successfully contacted, a brief description of the exchange is provided below.

In what follows each test is listed and designated by one of the variables to be compared. The actual test made or equation used comes on the next line. A brief discussion concludes the treatment.

As will be seen here, a very conservative approach was taken to making corrections to the data file. Some of the checks made were taken more as a way to look at the data and see if it fits preconceived notions that hard and fast rules had to be adhered to.[7]

The convention here is to list the question number (e.g., Q2) followed by what in context is a largely self-explanatory acronym, as in **Q2TRANS** -- the first item listed below. After the key question to be tested is the test itself, on a separate line. Some comments, then, follow indicating the disposition made or the results learned from recontacting respondents. Sometimes a clear "fix" was possible, in many cases nothing could or should be done, and, finally, a few inconsistencies were handled by making one of more of the inconsistent data items missing (to be imputed as described in Subsection I.F below).

[7] As a semi-theoretical note, if all consistency checks were enforced, further complicated by relationships assumed during imputation, then artificial paradoxes will be created even after many edits are made. The work at the Census Bureau on business data by Brian Greenberg may be consulted here. See, in particular, the Federal Committee on Statistical Methodology (FCSM) Report on Data Editing.

Q2TRANS
Q2TRANS x 365 <= upper_bound (Q3PROD)

This consistency check was rejected as a requirement. No editing was done. Note there was confusion whether <= or >= should be used – either way, too many companies fail the check, and the restriction does not seem to be necessarily true.

Q4RET
Q4RET+Q4CAT+Q4NET=100%

When this consistency check fails by more than 10%, all components are treated as missing. Otherwise, answers are left as is for the edited database. (Note: for the imputation, those answers left 'as is' will be scaled during the imputation – see Subsection I.E). When all "0"s are provided, "NA" is interpreted.

Q5TAXSL
Q5TAXSL <= Q5GROSS

Survey Unit (SUID 11030170) was called and the respondent indicated she made a mistake – Q5GROSS and Q5TAXSL were reversed – answers were manually corrected in NORCDatabaseOriginal. All other inconsistencies were edited, setting Q5GROSS to equal Q5TAXSL (Note: SUID 11017440 had the largest discrepancy – they could not be reached – they will be edited the same as the others).

Q6SHPTX
Q6SHPTX<= Q6USSHP;

Where this condition fails, Q6USSHP is set equal to Q6SHPTX

Q6USSHP
Q6USSHP<=Q5GROSS;

Q5SHPTX
Q6SHPTX<=Q5TAXSL;

Edits for these two consistency checks were not made. Suggested edits present two difficulties: (1) They might interfere with edits made due to other consistency failures, leading to potential artificial consistency failures; and (2) The recommended actions would alter a primary variable (a variable that is used in the tables and therefore, imputed) based on a non-primary variable (one not used in the tables). Such consistency checks would also complicate the imputation.

Q7USSHP
Q7USSHP<=Q7SHPTX;

Where this consistency check fails, Q7USSHP is set to Q7SHPTX

Q7USSHP
If Q6USSHP>0 then Q7USSHP>0;
If Q6SHPTX>0 then Q7SHPTX>0;

No consistency failures exist.

Q11CASH
Q11CASH+Q11CHKS+Q11DEBIT+Q11CARD+Q11OTHCR+Q11OTH=100%

When this consistency check fails by more than 10%, all components are treated as missing. Otherwise, answers are left as is for the edited database. When all "0"s are provided, "NA" is interpreted.

Q12DCARD
Q12DCARD<=5%; Q12CARD<=5%; Q12OTHCR<=5%

Each component is treated separately – when the condition for a component fails, that component is treated as missing.

Q12DCARD
If Q12DCARD > 0 then Q11DEBIT > 0;
If Q12CARD > 0 then Q11CARD > 0;
If Q12OTHCR > 0 then Q11OTHCR > 0;

No Consistency Failures Exist.

Q13CASHR
Q9CASHR >= Q13REG

This condition was rejected as a requirement. This is, by far, the most failed check – and to have more registers than cashiers does not seem like an inconsistency, anyways.

Q14MAN
Q14MAN+Q14SEMI+Q14AUTO=100%

For consistency failures where components add up to 1 through 17, a total of 27 cases, the sum represents the number of registers (Q13REG). Two other cases add up to 90% and 97% – all responses are all scaled to add up to 100%. Where all "0"s are written, it is assumed the respondent meant "NA" for all components (most have "0" written for Q13REG).

Q15MAN
Q15MAN < Q15SEMI < Q15AUTO;

A majority of failures for this consistency check exist where "0" is entered – it is simply assumed that "0" means "NA." No edits are made where non-zero answers are inconsistent with the consistency check.

Q15MAN
If Q15MAN>0 then Q14MAN>0;
If Q15SEMI>0 then Q14SEMI>0;
If Q15AUTO>0 then Q14AUTO>0;

No Consistency Failures Exist.

Q16REG
If Q13REG = 0 then Q16! = Yes

No Consistency Failures Exist.

Q25TOTAL
If Q25 and sum (Q17…Q24) both exist, then sum (Q17…Q24) is within $1 of Q25

PwC indicated one company failing this consistency check was to be contacted. The individual was both emailed and called but NORC still was unable to be reached.

For all cases where a discrepancy exists, the recommended solution by PwC was always to use the larger of sum (Q17…Q24) or Q25.

Q26PAY
Q26PAY <= Q25EST or Q26PAY <= sum (Q17 … Q24)

NORC attempted to contact all companies with discrepancies on this consistency check (9 consistency failures – 4 responses without contact information = 5 attempts). Three could not be contacted, the remaining two provided corrections to their answers. SUID 11015610 went from $2400 to $480, where Q25 is $1200 – SUID 10001700 from $500 to $0, where Q25 is $400. For the remaining 7 consistency failures, Q26 is marked as missing.

Q27TAXSL
If Q5TAXSL=0 then Q27STAX=0;
If Q5TAXSL=0 then Q27LTAX=0;

Q27 is not a variable used in the tables, thus it was not imputed. After review, it was felt an inconsistency here did not warrant altering Q5, which is an important variable. Therefore, nothing was done for the 6 inconsistencies here.

Q28STAX
Q27STAX>=Q28SJUR; Q27LTAX>=Q28LJUR;

This condition was tested as written, since PwC required it. Many failures exist here, no primary variables are affected, and it is not clear that the condition must be true.

Q28SJUR
If Q5TAXSL=0 then Q28SJUR=0;
If Q5TAXSL=0 then Q28LJUR=0;

There are five inconsistencies for this consistency check – all five also failed Q27cc. This is similar to Q27cc – no edits were made.

Q29REMIT
Q29REMIT <= 10% x Q5TAXSL

This consistency check had the highest number of failures behind Q13cc. NORC attempted to contact all 10 cases that PwC recommended calling:

- One company did not leave contact information, and so could not be reached.
- Two companies did not return NORC calls or emails
- Seven companies were contacted and ALL provided corrections to their responses:
    - SUID 11018550 changed their Taxable Sales from $135,686,000 to $1,985,000,000
    - SUID 11030390 changed their Remitted Tax from $5,000,000 to $1,240,000
    - SUID 11029990 changed their Remitted Tax from $213,462,447 to $2,134,624.47 AND changed their Taxable Sales from $284,584.02 to $56,111,583
    - SUID 11018750 changed their Taxable Sales from $3,000,000 to $30,000,000
    - SUID 13034500 indicated their Taxable Sales was not $393,766, but $393,766,000 – this error was made during NORC's keying – "In thousands" was written in the margin for Q5 – consequently, Q5GROSS has also been altered.
    - SUID 10002190 changed their Remitted Tax from $183,623,262 to $1,836,232.62
    - SUID 11016680 changed their Remitted Tax from $77,000,000 to $47,000,000

The above changes were rerun through consistency check Q5cc – no new consistency failures resulted. There is no pattern to the changes made – taxable sales is as equally likely to be changed as remitted tax. However, since Q5TAXSL is seen as a more important variable, it was decided to make Q29 missing where a consistency failure resulted. Also, the definition of a consistency check failure was expanded to 20% -- this allows 28 additional companies to 'pass' this consistency check.

So when a company reports taxable sales exceeding 20% of taxable sales, taxable sales is set as "missing." This nullifies answers to 27 companies. Q29 is a primary variable – therefore, this edit will likely have the largest effect on the final results. Remitted Tax for these 27 companies will now be imputed to a lower ratio of reported taxable sales.

Q30TAX
Q30TAX <= Q29REMIT

One company failed this check for which both variables are less than $10 – no edits were made.

In addition, three companies reported a large proportion of their sales tax as uncollected due to bad debt. Attempts to get these companies to clarify or double-check their responses were unsuccessful. Q30TAX are set to $0 for two cases and missing (-3) for the third case, because in the final analysis NORC and PwC believed the respondents misunderstood the question and provided incorrect answers on the survey.

Q31VEN
Q31VEN <= 5% x Q29REMIT

Seventeen (17) companies failed this check. PwC recommended we call four companies. One did not respond, and another refused to answer when called. The remaining two companies were contacted and insisted their answers were correct – one reported a 9.9% discount, and another a 15.9% discount. As a result, this condition was rejected as a requirement – except in one case where the company reported a 621% discount: Q31VEN is set to 'missing' for this case.

Q34NGOOD
Q34NGOOD+Q34CERT+Q34OUT+Q34EXMP+Q34OTH=100%

When this consistency check fails by more than 10%, all components are treated as missing. Otherwise, answers are left as is for the edited database. When all "0"s are provided, "NA" is interpreted.

Q35ELECT
Q35ELECT+Q35PAPER+Q35OTH=100%

Many companies indicated 100% for two columns, likely implying they store all records in two different ways. These are altered to 50% / 50%. Another indicated a sum that added to 150%, this is also scaled to add to 100%. When all "0"s are provided, "NA" is interpreted.

F. Filling in Data Gaps

Inconsistencies were noted in the way respondents filled out the questionnaires. In some cases, as noted above, these inconsistencies were "resolved" by setting one or more of the inconsistent items to missing.[8] Sometimes, though, no answer was given despite the fact one was required.

[8] The basic data file, as documented in Section II, from the survey makes it straightforward to look at such cases, since the edited and imputed questionnaire records have been concatenated and either version can be used.

In both these instances, if the question was to be used in the summary tabulations shown in Section V of Volume I of this report, then a hot deck imputation was made and the entry filled in for tabular use (following Fellegi and Holt (1976)). Here we describe this "fill in" or imputation process briefly.

In general, a missing value is not imputed if the respondent marked "NA" (-5). This was considered to be an answer and was interpreted as '0' for tabulations. However, in other cases, for the key variables used in the basic tables, a form of imputation was employed, if the item was blank, to fill-in items where possible. Some of the imputations were done manually; most were made with a hot deck (e.g., Ford 1983) where in the main the cell definitions were D&B Sales and the state sales tax grouping or "state bucket." The variables that defined the hot deck were an exception to this general rule and had to be manually imputed.

F.1 Manually Imputed Variables. The "frame" variables (those items from the D&B database or created during the design stage) were all manually imputed.

SUID

For 18 companies the questionnaire was returned without an identifying Survey Unit Identifier (SUID). This can occur (a) if the company did not wish to be identified, so they did not return the cover page, (b) if the company downloaded the survey off the web, or (c) if the company was not selected in the original design. In these 18 instances we proceeded as follows.

DBSIC

The SIC code exists for all but the 18 companies not matched to the frame. For these 18, the 4-digit SIC was simply derived from the survey answers to question 39. All 18 businesses provided an answer to this survey question, although two did not provide any detail beyond simply identifying themselves as a retail store. As such, they were coded 59990000: General Retail.

Statebucket or DB State

These details exist for all but the 18 companies not matched to the frame. Survey answers to question 8 were used to fill in values, where possible. In 15 of the 18 missing cases, respondents answered Question 8 and all indicated their stores were located in only one state. That state was used to fill in missing values. For two cases, question 8 was left unanswered, but the area code of the phone number provided as a contact (not available in the PwC database) was used to pinpoint the location of the business. The final missing case has no information to identify the

company's location. The survey answers indicate they pay sales tax, so they were simply placed in the largest sales-tax-paying state bucket, which turned out to be state bucket 4.[9]

F.2 Hot Deck Imputed Variables. The questionnaire variables to be used in the summary tabulations shown in Section V of Volume I of this report were all imputed using a hot deck.

Q5GROSS
Where Q5TAXSL exists, it is used as the donor amount – otherwise, D&B Sales is used as the donor amount. Where neither exists, the case is unusable.

128 missing: 9 marked 'NA', 116 imputed, 3 remained unchanged.

Q5TAXSL
Donor must have the same SIC and must have Q5TAXSL <= Q5GROSS – closest donor on Q5GROSS (capped at double the missing case's total sales plus $1) is used to impute Q5TAXSL proportionally as follows:

> Q5TAXSL_imp = Q5GROSS_imp * (Q5TAXSL_donor / Q5GROSS_donor)
> All others are simply set to Q5GROSS where this variable exists.
>
> The exception is state bucket 9 – where a company has stores only in state bucket 9, and no remote sales are reported, Q5TAXSL_imp is set to 0.

As a consequence of discussions with PwC, it was decided to 'edit out,' just for the imputed database, those companies in state bucket 9 who reported positive taxable sales with no remote sales. Here, as well, Q5TAXSL is set to 0. Where remote sales do exist, the percent of remote sales is applied to the ratio obtained from the donor company to impute taxable sales.

126 missing: 13 marked 'NA', 110 imputed, 3 remained unchanged

[9] It should be mentioned here that these 'imputations' were made to the original database as well.

In the above figure the Y-axis represents Proportion of Taxable Sales to Gross Sales for companies answering both (X-axis is simply all companies in ascending order of this proportion). Small plateau at right is "Proportion = 1." Cases that originally extended beyond the top of the graph, indicating reported taxable sales that exceeded gross sales, were altered during the editing stage: Q5TAXSL was set to missing.

Q29REMIT

Donor must be in same state bucket and be in the same general industry (general industry is the first two digits of the SIC code) – closest donor on Q5TAXSL (capped at double the missing case's taxable sales plus $1) is used to impute proportionally:

> Q29REMIT_imp = Q5TAXSL_imp * (Q29REMIT_donor / Q5TAXSL_donor)
> Only donors where Q29REMIT <= 10% x Q5TAXSL are used

The same edit as discussed in Q5TAXSL is applied here: companies in state bucket 9 without remote sales are given Q29REMIT_imp = 0 even if Q29REMIT is positive.

Following further discussions with PwC, it was decided SUID 13039970 was an outlier and should not be used as a donor here.

151 missing: 9 marked 'NA', 131 imputed, 11 remained unchanged.

Q8DISTINCT

This is a derived variable – it represents the number of tax-paying states (including DC) in which the company pays sales tax. This is derived from the many cells in question 8 of the survey.

> Where the cells are empty, a '0' was entered during keying. Where all cells are '0,' the question is considered missing, and Q8DISTINCT is imputed.

The simplest imputing strategy was used: all companies are assumed to have at least one store – and in fact, most responding companies have only one store. So Q8DISTINCT is set to 1 where missing, except where STATEBUCKET indicates the store does not pay sales tax – in the latter case, Q8DISTINCT is set to 0. If, as in a few cases, question 8 is missing but Q28SJUR exists, Q8DISTINCT is simply set equal to Q28SJUR.

74 missing: 74 imputed

Q9TOTAL
Question 9 has 5 parts for the respondent to indicate the number of employees for various categories of employees. For imputation, only the total is imputed where missing. Furthermore, the imputation is done in the simplest manner possible: if any part of question 9 was answered, it is assumed that the unanswered parts are zero. Q9TOTAL is then simply the sum of the non-missing parts.

For companies with D&B sales of a $1M+, the D&B employee field is used where Q9TOTAL is missing. For companies with D&B sales below $1M, the donor company must match the first two digits of the SIC – the closest donor on sales is used directly to impute Q9TOTAL:

Q9TOTAL_imp = Q9TOTAL_donor.

25 missing: 24 imputed, 1 remained unchanged.

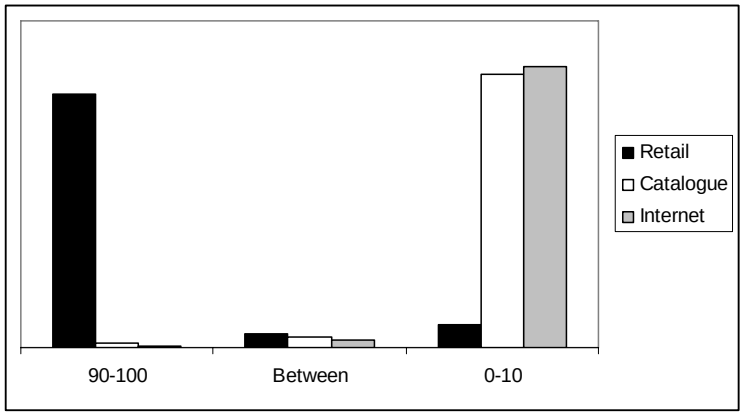
Q3PROD
Donor must have the same SIC – closest donor on Q5GROSS (capped at double the missing case's gross sales plus $1) is used:

Q3PROD_imp = Q3PROD_donor

37 missing: 2 marked 'NA', 33 imputed, 2 remained unchanged.

Q4RET / Q4CAT / Q4NET
As a result of the editing, all responses are in 'clean' categories of either all "NA", all "missing", or numbers that sum to 100% (or close) with no missing components. Four companies supplied numbers that did not add to 100% but came close: for these, values were manually altered in a manner consistent with the supplied values so that they added to 100%. Non-missing values occur at the frequencies shown in the graph. The most popular survey answer by far for Q4 is "100% Retail." Therefore, missing values are imputed to be "100% Retail" where Q6USSHP is zero or missing. In the (only three) cases where Q6USSHP exists and is greater than zero, the proportion Q6USSHP/Q5GROSS is divided equally between Q4CAT and Q4NET with Q4RET getting 100% minus that proportion.

37 missing: 18 marked 'NA', 19 imputed.



Relative frequency of percentage ranges for non-missing Q4. This graph supports the method taken of imputing used for most missing cases.

COMPLIANCE COSTS: Q17 – Q25

As a result of the editing, Q17 through Q25 are "clean" – that is, where they are nonmissing, Q17 … Q24, then the sum add up to Q25. For missing responses, all components are cleanly marked as missing or "NA".

The imputation here was limited to Q25EST, no attempt was made to determine the individual components where compliance costs were left blank. Where Q29REMIT is given as "0" or "NA", then compliance costs are imputed to be "0" as long as it is missing (if a company reported no sales tax but positive compliance costs, no attempt was made to 'correct' the response). Otherwise, Q25EST is imputed based upon a donor, which must be in the same general SIC and state bucket – the closest donor on Q5GROSS (capped at double the missing case's sales plus $1) is used to impute proportionally:

Q25EST_imp = Q5GROSS_imp * (Q25EST_donor / Q5GROSS_donor)

An exception to this imputation is where Q26PAY > Q25EST_imp – in this case, Q25EST_imp is set to Q26PAY. It is assumed compliance costs for these companies are equal to what they pay an outside contractor.

A second exception follows from edits in Q5TAXSL_imp and Q29REMIT_imp for state bucket 9: if a company is completely in state bucket 9 and reports no remote sales, then Q25EST_imp is set to 0, even if Q25EST is greater than 0. In these cases, Q17-Q24 are also set to 0, as are other related items such as Q26PAY and Q31VEN.

A final exception is that SUID 12015980 cannot be used as a donor since this company was found to be an outlier following discussions with PwC.

137 missing: 15 marked 'NA', 117 imputed, 5 remain unchanged.

Q30TAX
Donor must be in same state bucket and general SIC – closest donor on Q5GROSS (capped at double the missing case's gross sales plus $1) is used to impute proportionally:

Q30TAX_imp = Q5GROSS_imp * (Q30TAX_donor / Q5GROSS_donor)

Cases without donors were set to 0. If imputation causes Q30TAX to exceed Q29REMIT, then Q30TAX is set equal to Q29REMIT.

195 missing: 28 marked 'NA', 153 imputed, 14 set to 0.

Q26PAY
Donor must be in same state bucket and general SIC – closest donor on Q25EST (capped at double the missing case's total compliance costs plus $1) is used to impute proportionally:

Q26PAY_imp = Q25EST_imp * (Q26PAY_donor / Q25EST_donor)

Cases without donors were set to 0.

217 missing: 23 marked 'NA', 180 imputed, 14 set to 0.

Q31VEN
Donor must be in same state bucket – closest donor on Q29REMIT (capped at double the missing case's total sales tax remitted plus $1) is used to impute proportionally:

Q31VEN_imp = Q29REMIT_imp * (Q31VEN_donor / Q29REMIT_donor)

Cases without donors were set to 0.

168 missing: 42 marked 'NA', 111 imputed, 15 set to 0.

CREDITFEES
This is a derived variable:

CREDITFEES = Q29REMIT * (Q11DEBIT*Q12DCARD +
Q11CARD*Q12CARD + Q11OTHRCR*Q12OTHRCR)

Answers to question 11 and 12 are first cleaned. Answers to question 11 are made to add to 100 if the sum lies within 90 to 110 – otherwise CREDITFEES is marked as missing.

Missing question 12 answers are set to 0 if at least one non-zero component exists – otherwise CREDITFEES is marked as missing. If -5 ("NA") is marked for all of Q11 or all of Q12, then CREDITFEES is set to '0'.

If Q29REMIT as well as questions 11 and 12 are non-missing, then CREDITFEES are calculated as above and act as donors for all other cases. Closest donor on Q29REMIT is used to impute proportionally:

CREDITFEES_imp = Q29REMIT_imp * (CREDITFEES_donor / Q29REMIT_donor)

Cases without donors are set to 0.

208 missing: 180 imputed, 28 set to 0.

POSFLOAT
This is a derived variable:

POSFLOAT = Q29REMIT*Q32REMIT*Q32A/365*4.12%

Where Q29REMIT, Q32REMIT, and Q32A exist, POSFLOAT is calculated as above and can be used as a donor. Donor must be in same general SIC and state bucket. The closest donor on Q29REMIT is used to impute proportionally:

POSFLOAT_imp = Q29REMIT_imp*(POSFLOAT_donor/Q29REMIT_donor)

Cases without donors are set to 0.

113 missing: 27 marked 'NA', 83 imputed, 3 set to 0.

NEGFLOAT
This is a derived variable:

NEGFLOAT = Q29REMIT*Q32REMIT*Q32A/365*4.12%

Where Q29REMIT, Q32REMIT, and Q32A exist, NEGFLOAT is calculated as above and can be used as a donor. Donor must be in same general SIC and state bucket. The closest donor on Q29REMIT is used to impute proportionally:

NEGFLOAT_imp = Q29REMIT_imp*(NEGFLOAT_donor/Q29REMIT_donor)

Cases without donors are set to 0. Also, all cases where a negative float was imputed even though the company reported 100% of their sales occurring before sales tax remittance were set to 0.

150 missing: 30 marked 'NA', 115 imputed, 5 set to 0.

G. Survey Estimation

There were 796 survey questionnaires completed for the study, roughly 8% of those mailed that turned out to be in business and eligible. Only the 762 of these questionnaires were brought forward to the estimation or weighting step and included in the summary analytic tables to be found in Section V of Volume I of this report. The 34 questionnaires omitted were judged to be just too incomplete to use.

A great deal of work was done in determining how to make the final estimates. The frame was known to be somewhat out of date and to have other problems (See Subsection I.D above). We also had the nonresponse to deal with. What we ended up doing was to ratio adjust the respondents shown in Table 6 (excluding the 34 that were judged unusable) up to the original D&B frame marginal totals, arguing in effect that the firms no longer in business basically equaled those that were still on the frame because it was out-of-date. We felt comfortable in doing this because the Census Bureau totals for the retail sector were so close to what D&B gave us (both around 630,000).

To deal with the problem of cases where the D&B sales amount and that reported on the survey were in different classes we always used the larger of the two. This was a way to protect us against having sampled a case in too low a size class, because of problems in inconsistent treatment of enterprises and establishments in the D&B frame. Table 7 provides the final weights employed in each state group and sales size class.

The top 50 companies, which were given special attention, were all given a weight of 3.125 with the exception of the largest company, which was given a weight of 1. One case could not be matched to the frame: its reported annual retail sales were less than $150,000, so it was included in the results with a weight of 1.

Bottom line here: The final estimates for businesses and also for number of employees obtained in the State Sales Tax Compliance Study turn out, after weighting up the survey to be reasonably close, so no additional efforts were made to force them to Census or Bureau of Labor Statistics totals.

Table 7. Final Weights (rounded)

Sales Class	State Bucket								
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Under $150k	-	-	-	1	-	-	-	-	-
150k - 200k	9,274	6,817	7,305	-	5,002	17,029	2,442	7,140	-
200k - 500k	5,305	2,283	3,547	2,862	7,387	4,792	3,727	2,892	688
500k – 1M	3,406	1,893	4,928	1,903	1,994	2,503	10,081	-	459
1M - 2.5M	821	1,025	628	805	1,007	1,311	1,098	1,742	173
2.5M – 5M	316	560	351	205	431	294	2,279	603	94
5M – 10M	280	1,147	653	172	205	208	179	248	47
10M – 25M	140	75	89	114	195	130	341	120	25
25M – 50M	58	184	44	44	93	54	35	59	34
50M - 100M	44	29	24	29	19	54	49	20	55
100M+	14	20	16	19	15	34	19	15	38

H. Conclusions and Recommendations

As was mentioned at the outset, the focus of the NORC documentation is not simply to describe what was done -- albeit that is clearly our first goal. Rather we have looked at how well the Study resources in time, money and opportunity costs were used to handle the four major challenges or errors that confront all survey work: coverage error, missing data, measurement error and sampling error. It is now time to sum up our views and to make recommendations on any unresolved analytic issues these data might have.

H.1 General. Of the four error sources only the impact of sampling errors can be quantified fully. As detailed in Section IV of this Volume, NORC computed the sampling variance using the random group method (Wolter 1985).

Coverage errors depend on the frame or list that is available for use or that can be constructed. We believe, as noted in Subsection I.G, that some optimism may be warranted. More might be done in comparing Census and other data sources to confirm this, however.

Measurement errors depend on how careful respondents are and how well they understand what is wanted. This has been covered in enough detail already to give PwC a way to successfully

interpret the data obtained. We feel, again, that some optimism is warranted, at least to the extent that the errors made are modest enough relative to sampling error, so that when sampling variances are calculated the margins of error obtained will reflect the error from this source fairly well.

Of course, the last concern, that of missing data, can be very harmful and has too in many cases, including in this survey, largely only be speculated about. We did do a special study here, however, and found that most of the unit or complete noninterview nonresponse is ignorable and not biasing (i.e., this means that the penalty is confined just to a smaller than desired sample size). Information about the biases arising from the item nonresponse has not been assembled,[10] but we have recommended that sensitivity analyses be conducted. Certainly, if an item is of importance and often missing, a second imputation might be employed using more subject matter knowledge than the NORC survey team and its hot deck approach could bring to bear.

H.2 Microsimulation Use of Study. One final specific observation may be worth making about the value of the Study and its limitations. Here the context is within a microsimulation modeling effort where two or more alternatives are "scored" against each other.

The extensive assessment we have made of the Study results gives us some comfort that the data can be relied upon to robustly rank such "What If" alternatives. Harder to be convinced of is the direct value of any level estimates that might result. Percentage results and dollar ratios will usually be sounder that overall totals, unless these totals are anchored on information from outside of the survey.

H.3 Bottom Line. In summary, then, it is our view that, if employed properly, this first-ever National Study of state sales tax compliance costs is "fit for the uses" for which it was intended. This is so, despite important limitations that we have treated in detail here.

[10] The technique used is described in Scheuren (2005a, 2005b).

References

Cochran, W. (1977). ***Sampling Techniques***. Wiley: New York.

Deming, W.E. (2006) On errors in surveys. ***The American Statistician.*** An excerpt republished from the 1943 ***American Sociological review***

Federal Committee on Statistical Methodology (FCSM) Report on Data Editing (1993). www.FCSM.gov

Fellegi, I. P. and D. Holt. 1976. A systematic approach to automatic edit and imputation. ***Journal of the American Statistical Association*** 71: 17-35.

Ford, B. M. 1983. An overview of hot-deck procedures. In ***Incomplete Data in Sample Surveys, vol. 2, Theory and bibliographies,*** edited by W. Madow, I. Olkin and D. Rubin. New York: Academic Press.

Scheuren, F. (2005a). Paradata from concept to completion. Presented at the 2005 Statistics Canada Methodology Conference

Scheuren, F. (2005b). November History Corner ***The American Statistician .***

Wolter, K. (1985). ***Introduction to Variance Estimation***. Springer-Verlag.

II. CODEBOOK COUNTS

Included here are two sets of counts:

Part I contains the number and percent of each usable response by question, after imputation and editing

Part II contains the number and percent all responses by question, after editing but before imputation

There are counts for 762 complete questionnaires in Part I. In Part II the total count is 34 larger. Part II includes questionnaires that were in the end judged to be too incomplete to be usable.

Only the 762 questionnaires in Part I were brought forward to the weighting step and included in the summary analytic tables in Section V of Volume I of this report. Information on the techniques used to edit, impute and weight the survey data are detailed in Section I of NORC's documentation of its work on the survey.

Both parts are similar in structure. The questions are first stated in their entirety. If the question called for an amount or a percent then the valid values are tabulated by size. If preset categories were to be checked these are shown as they appeared on the questionnaire.

Opposite each question are to be found the number (denoted by "n") and percent (denoted by "pct") of times that a particular response was given in the sample. Not all the questions, of course, needed to be answered and so in many cases no answer may have been provided for an item.

Sometimes, though, no answer was given despite the fact one was required. If the question was to be used in the summary tabulations shown in Section V of Volume I of this report, then a hot deck imputation was made and the entry filled in for tabular use. The imputed counts found in Part I show the item values after imputation. In Part II the counts are shown before imputation.

By comparing Part I (where some data items are imputed) and Part II (where no items have been imputed), it can be seen that the number of items imputed varied considerably. We cannot offer any rule of thumb here but do recommend that if the item is of importance and often missing, a second imputation might be employed using more subject matter knowledge than the NORC survey team could bring to bear. Items that have very few, or zero, unanswered responses after imputation are generally those that were imputed.

Take, for example –

Q3: How many different products (e.g., SKUs)
did you sell at retail as of year-end 2003?

After imputation there were no unanswered questionnaires as tabulated in Part I. In Part II there were 37 unanswered cases to this question that had to be imputed.

The basic data file from the survey makes it easy to look at such cases, since the edited and imputed questionnaire records have been concatenated and either version can be used. On the records, too, are found a user id number, the sample weight, plus information from the Dun and Bradstreet frame used in making the initial selections.

Part I – Fully Edited and Imputed Survey Data by Question
(JCCS Sample of 762 Usable Observations)

Q1: What do you consider the greatest cost in collecting and remitting sales tax?

Category	*n*	*pct*
Training personnel	36	5%
Programming cash register/POS systems	50	7%
Preparing tax forms including research	327	43%
Remitting sales tax	64	8%
Handling audits	45	6%
Keeping track of local taxes	44	6%
Documenting exempt sales	85	11%
Other	42	6%
Unanswered	69	9%

Q2: About how many retail sales transactions (invoices) did you have per day in 2003?

Average number of daily retail sales transactions	*n*	*pct*
0 to 10	62	8%
10 to 100	270	35%
100 to 1000	224	29%
1000 to 10,000	74	10%
10,000 plus	52	7%
Unanswered	80	11%

Q3: How many different products (e.g., SKUs) did you sell at retail as of year-end 2003?

Category	*n*	*pct*
Less than 1,000	296	39%
1,000 to 5,000	160	21%
5,000 to 10,000	92	12%
10,000 to 25,000	88	12%
25,000 to 50,000	52	7%
50,000 to 100,000	39	5%
100,000 or more	35	5%
Unanswered	0	0%

Q4: What percent of your retail sales dollars in 2003 were through the following channels?

Retail store sales	*n*	*pct*
0% to 10%	23	3%
11% to 20%	0	0%
21% to 30%	6	1%
31% to 40%	3	0%
41% to 50%	2	0%
51% to 60%	3	0%
61% to 70%	3	0%
71% to 80%	14	2%
81% to 90%	27	4%
91% to 100%	681	89%
Unanswered	0	0%

Catalogue sales	*n*	*pct*
0% to 10%	716	94%
11% to 20%	9	1%
21% to 30%	7	1%
31% to 40%	3	0%
41% to 50%	3	0%
51% to 60%	2	0%
61% to 70%	4	1%
71% to 80%	3	0%
81% to 90%	5	1%
91% to 100%	10	1%
Unanswered	0	0%

Internet sales	*n*	*pct*
0% to 10%	738	97%
11% to 20%	10	1%
21% to 30%	9	1%
31% to 40%	0	0%
41% to 50%	0	0%
51% to 60%	0	0%
61% to 70%	1	0%
71% to 80%	0	0%
81% to 90%	0	0%
91% to 100%	4	1%
Unanswered	0	0%

Q5: Please provide the following information for your U.S. retail activities in 2003:		
Gross sales before returns and allowances	*n*	*pct*
Under $150k	31	4%
$150k to $200k	9	1%
$200k to $500k	75	10%
$500k to $1M	53	7%
$1M to $2.5M	105	14%
$2.5M to $5M	74	10%
$5M to $10M	68	9%
$10M to $25M	102	13%
$25M to $50M	79	10%
$50M to $100M	66	9%
$100M plus	93	12%
Unanswered	7	1%
Taxable sales	*n*	*pct*
Under $150k	128	17%
$150k to $200k	23	3%
$200k to $500k	78	10%
$500k to $1M	51	7%
$1M to $2.5M	83	11%
$2.5M to $5M	84	11%
$5M to $10M	69	9%
$10M to $25M	76	10%
$25M to $50M	56	7%
$50M to $100M	43	6%
$100M plus	62	8%
Unanswered	9	1%
Q6: How much were your remote sales (catalogue or Internet), if any, in 2003?		
Shipments to all US locations	*n*	*pct*
NA: $0 Entered	449	59%
Under $150k	42	6%
$150k to $200k	3	0%
$200k to $500k	5	1%
$500k to $1M	6	1%
$1M to $2.5M	7	1%
$2.5M to $5M	7	1%
$5M to $10M	4	1%
$10M to $25M	14	2%
$25M to $50M	6	1%
$50M to $100M	2	0%
$100M plus	15	2%
Unanswered	202	27%

Shipments on which you collect and remit sales tax	*n*	*pct*
NA: $0 Entered	443	58%
Under $150k	38	5%
$150k to $200k	1	0%
$200k to $500k	4	1%
$500k to $1M	5	1%
$1M to $2.5M	6	1%
$2.5M to $5M	5	1%
$5M to $10M	7	1%
$10M to $25M	11	1%
$25M to $50M	1	0%
$50M to $100M	1	0%
$100M plus	12	2%
Unanswered	228	30%

Q7: How many states (including the District of Columbia) did you ship to in 2003?

Number of states shipped to	*n*	*pct*
Zero	256	34%
One	143	19%
Two	44	6%
Three to Five	84	11%
Six t Nine	23	3%
Ten to Twenty	41	5%
21 to 40	15	2%
41 and up	45	6%
Unanswered	111	15%

Number of states shipped to for which you collect and remit sales tax	*n*	*pct*
Zero	313	41%
One	175	23%
Two	27	4%
Three to Five	41	5%
Six t Nine	13	2%
Ten to Twenty	5	1%
21 to 40	16	2%
41 and up	16	2%
Unanswered	156	20%

Q8: Please indicate below the number of retail stores you had in each state (including the District of Columbia), if any, as of June 30, 2003:

State	*n*	*pct*
AL:	649	2%
AK:	158	0%
AZ:	841	2%
AR:	435	1%
CA:	4,144	10%
CO:	620	2%
CT:	468	1%
DE:	126	0%
DC:	41	0%
FL:	3,039	7%
GA:	1,190	3%
HI:	175	0%
ID:	234	1%
IL:	2,244	6%
IN:	1,054	3%
IA:	472	1%
KS:	390	1%
KY:	599	1%
LA:	677	2%
ME:	149	0%
MD:	619	2%
MA:	907	2%
MI:	1,457	4%
MN:	672	2%
MS:	511	1%
MO:	856	2%
MT:	182	0%
NE:	266	1%
NV:	390	1%
NH:	257	1%
NJ:	1,082	3%
NM:	278	1%
NY:	1,879	5%
NC:	1,081	3%
ND:	115	0%
OH:	1,846	5%
OK:	450	1%
OR:	454	1%
PA:	1,443	4%
RI:	143	0%
SC:	520	1%
SD:	107	0%
TN:	883	2%
TX:	3,367	8%
UT:	393	1%
VT:	75	0%
VA:	894	2%

WA:	856	2%
WV:	180	0%
WI:	835	2%
WY:	81	0%

Q9: How many employees did you have at the end of 2003?

Employees in tax department (fraction if less than 1)	*n*	*pct*
Zero	123	16%
Fraction under 1	126	17%
1 to 2.9	242	32%
3 to 5	33	4%
6 to 9	9	1%
10 to 20	19	2%
21 to 40	10	1%
41 or higher	17	2%
Unanswered	183	24%

Employees in accounting department (fraction if less than 1)	*n*	*pct*
Zero	44	6%
Fraction under 1	61	8%
1 to 2.9	238	31%
3 to 5	163	21%
6 to 9	66	9%
10 to 20	48	6%
21 to 40	18	2%
41 or higher	29	4%
Unanswered	95	12%

Employees in customer service department (fraction if less than 1)	*n*	*pct*
Zero	70	9%
Fraction under 1	18	2%
1 to 2.9	177	23%
3 to 5	118	15%
6 to 9	61	8%
10 to 20	70	9%
21 to 40	39	5%
41 or higher	57	7%
Unanswered	152	20%

Cashiers		
(fraction if less than 1)	*n*	*pct*
Zero	71	9%
Fraction under 1	23	3%
1 to 2.9	225	30%
3 to 5	104	14%
6 to 9	37	5%
10 to 20	48	6%
21 to 40	27	4%
41 or higher	88	12%
Unanswered	139	18%

Other employees		
(fraction if less than 1)	*n*	*pct*
Zero	38	5%
Fraction under 1	5	1%
1 to 2.9	73	10%
3 to 5	51	7%
6 to 9	38	5%
10 to 20	84	11%
21 to 40	88	12%
41 or higher	244	32%
Unanswered	141	19%

Q10: About what percent of the gross sales of your retail business were:

Returned or exchanged	*n*	*pct*
NA: 0%	149	20%
Under 10%	462	61%
10% to 20%	19	2%
20% to 30%	4	1%
30% to 40%	2	0%
40% to 50%	0	0%
50% to 60%	0	0%
60% to 70%	0	0%
70% to 80%	1	0%
80% to 90%	0	0%
90% to 100%	1	0%
Unanswered	124	16%

Written off as a bad debt	*n*	*pct*
NA: 0%	221	29%
Under 10%	397	52%

10% to 20%	5	1%
20% to 30%	2	0%
30% to 40%	0	0%
40% to 50%	0	0%
50% to 60%	0	0%
60% to 70%	0	0%
70% to 80%	0	0%
80% to 90%	0	0%
90% to 100%	0	0%
Unanswered	137	18%

Q11: Approximately what percent of your total sales dollars were paid in the following ways in 2003?

Cash	*n*	*pct*
0% to 10%	309	41%
10% to 20%	122	16%
20% to 30%	88	12%
30% to 40%	50	7%
40% to 50%	37	5%
50% to 60%	19	2%
60% to 70%	19	2%
70% to 80%	16	2%
80% to 90%	8	1%
90% to 100%	8	1%
Unanswered	86	11%

Checks	*n*	*pct*
0% to 10%	166	22%
10% to 20%	103	14%
20% to 30%	87	11%
30% to 40%	64	8%
40% to 50%	64	8%
50% to 60%	32	4%
60% to 70%	25	3%
70% to 80%	52	7%
80% to 90%	32	4%
90% to 100%	51	7%
Unanswered	86	11%

Debit cards	*n*	*pct*
0% to 10%	570	75%
10% to 20%	66	9%
20% to 30%	29	4%
30% to 40%	7	1%
40% to 50%	0	0%

50% to 60%	1	0%
60% to 70%	2	0%
70% to 80%	0	0%
80% to 90%	1	0%
90% to 100%	0	0%
Unanswered	86	11%

In-House credit cards	*n*	*pct*
0% to 10%	623	82%
10% to 20%	21	3%
20% to 30%	9	1%
30% to 40%	7	1%
40% to 50%	6	1%
50% to 60%	4	1%
60% to 70%	1	0%
70% to 80%	3	0%
80% to 90%	1	0%
90% to 100%	1	0%
Unanswered	86	11%

Other credit cards	*n*	*pct*
0% to 10%	271	36%
10% to 20%	125	16%
20% to 30%	92	12%
30% to 40%	59	8%
40% to 50%	67	9%
50% to 60%	24	3%
60% to 70%	17	2%
70% to 80%	13	2%
80% to 90%	5	1%
90% to 100%	3	0%
Unanswered	86	11%

Other	*n*	*pct*
0% to 10%	571	75%
10% to 20%	16	2%
20% to 30%	12	2%
30% to 40%	5	1%
40% to 50%	9	1%
50% to 60%	14	2%
60% to 70%	19	2%
70% to 80%	15	2%
80% to 90%	9	1%
90% to 100%	6	1%
Unanswered	86	11%

Q12: For each of the following types of payment, indicate the average percentage fee you paid to the credit card company or other financial institution in 2003:

Fee for debit cards	*n*	*pct*
0% to .5%	112	15%
.5% to 1%	22	3%
1% to 1.5%	50	7%
1.5% to 2%	65	9%
2% to 3%	119	16%
3% to 4%	50	7%
4% to 5%	6	1%
5% to 10%	7	1%
10% to 20%	0	0%
20% to 100%	0	0%
Unanswered	331	43%

Fee for in-house credits cards	*n*	*pct*
0% to .5%	182	24%
.5% to 1%	1	0%
1% to 1.5%	4	1%
1.5% to 2%	17	2%
2% to 3%	39	5%
3% to 4%	11	1%
4% to 5%	3	0%
5% to 10%	1	0%
10% to 20%	0	0%
20% to 100%	0	0%
Unanswered	504	66%

Fee for other credit cards	*n*	*pct*
0% to .5%	55	7%
.5% to 1%	3	0%
1% to 1.5%	20	3%
1.5% to 2%	120	16%
2% to 3%	257	34%
3% to 4%	121	16%
4% to 5%	13	2%
5% to 10%	6	1%
10% to 20%	0	0%
20% to 100%	0	0%
Unanswered	167	22%

Q13: How many cash registers (including POS terminals and cash box/calculators) did you use in 2003?

Number of cash registers	*n*	*pct*
Zero	53	7%
One	227	30%
Two	127	17%
Three to Five	131	17%
Six to Nine	43	6%
Ten to Twenty	48	6%
21 to 50	20	3%
50 and up	80	11%
Unanswered	33	4%

Q14: About what percent of all cash registers used by your employees were of the following types in 2003?

Manual (including cash box and calculator):

Percentage Range	*n*	*pct*
0% to 10%	411	54%
10% to 20%	4	1%
20% to 30%	2	0%
30% to 40%	8	1%
40% to 50%	10	1%
50% to 60%	2	0%
60% to 70%	0	0%
70% to 80%	0	0%
80% to 90%	1	0%
90% to 100%	251	33%
Unanswered	73	10%

Semi-manual (without electronic data files):

Percentage Range	*n*	*pct*
0% to 10%	562	74%
10% to 20%	5	1%
20% to 30%	0	0%
30% to 40%	3	0%
40% to 50%	7	1%
50% to 60%	0	0%
60% to 70%	1	0%
70% to 80%	2	0%
80% to 90%	1	0%
90% to 100%	108	14%
Unanswered	73	10%

Automatic registers with electronic data files:		
Percentage Range	*n*	*pct*
0% to 10%	371	49%
10% to 20%	1	0%
20% to 30%	0	0%
30% to 40%	1	0%
40% to 50%	5	1%
50% to 60%	0	0%
60% to 70%	7	1%
70% to 80%	7	1%
80% to 90%	4	1%
90% to 100%	293	38%
Unanswered	73	10%

Q15: What was the approximate cost for a new cash register of each type that you used in your retail business in 2003?

Manual	*n*	*pct*
NA	0	0%
$0 to $5	1	0%
$5 to $20	8	1%
$20 to $50	16	2%
$50 to $100	39	5%
$100 to $200	42	6%
$200 to $500	50	7%
$500 to $1000	18	2%
$1000 to $2000	6	1%
$2000 to $5000	3	0%
$5000 and up	1	0%
Unanswered	578	76%

Semi-manual	*n*	*pct*
NA	0	0%
$0 to $5	0	0%
$5 to $20	0	0%
$20 to $50	1	0%
$50 to $100	2	0%
$100 to $200	13	2%
$200 to $500	29	4%
$500 to $1000	31	4%
$1000 to $2000	9	1%
$2000 to $5000	13	2%
$5000 and up	4	1%
Unanswered	660	87%

Automatic	*n*	*pct*
NA	0	0%
$0 to $5	1	0%
$5 to $20	0	0%
$20 to $50	2	0%
$50 to $100	1	0%
$100 to $200	4	1%
$200 to $500	20	3%
$500 to $1000	39	5%
$1000 to $2000	63	8%
$2000 to $5000	77	10%
$5000 to $10000	39	5%
$10000 and up	31	4%
Unanswered	485	64%

Q16: Would you have purchased less expensive cash registers if there were no sales tax?

Answer	*n*	*pct*
Yes	71	9%
No	559	73%
Unanswered	132	17%

Q16a: If yes, approximately how much more do you pay for a typical cash register than you would have paid in the absence of any sales tax?

Category	*n*	*pct*
NA	691	91%
$0 to $5	0	0%
$5 to $20	0	0%
$20 to $50	1	0%
$50 to $100	3	0%
$100 to $200	6	1%
$200 to $500	15	2%
$500 and up	31	4%
Unanswered	15	2%

Q17: Sales Tax Compliance Costs: Training Personnel on sales tax

Category	*n*	*pct*
$0 (none)	117	15%
Under $50	3	0%

$50 to $100	11	1%
$100 to $500	97	13%
$500 to $1k	42	6%
$1k to $5k	96	13%
$5k to $10k	26	3%
$10k to $100k	36	5%
$100k to $500k	1	0%
$500k plus	4	1%
Unanswered	329	43%

Q18: Sales Tax Compliance Costs:
Documenting tax-exempt sales

Category	*n*	*pct*
$0 (none)	72	9%
Under $50	13	2%
$50 to $100	9	1%
$100 to $500	107	14%
$500 to $1k	40	5%
$1k to $5k	106	14%
$5k to $10k	32	4%
$10k to $100k	45	6%
$100k to $500k	3	0%
$500k plus	6	1%
Unanswered	329	43%

Q19: Sales Tax Compliance Costs:
Customer service relating to sales tax issues
other than documenting exempt sales

Category	*n*	*pct*
$0 (none)	200	26%
Under $50	7	1%
$50 to $100	11	1%
$100 to $500	70	9%
$500 to $1k	28	4%
$1k to $5k	74	10%
$5k to $10k	19	2%
$10k to $100k	18	2%
$100k to $500k	3	0%
$500k plus	3	0%
Unanswered	329	43%

Q20: Sales Tax Compliance Costs:
Sales tax-related software acquisitions and license fees

Category	*n*	*pct*
$0 (none)	220	29%
Under $50	4	1%
$50 to $100	7	1%
$100 to $500	62	8%
$500 to $1k	32	4%
$1k to $5k	64	8%
$5k to $10k	7	1%
$10k to $100k	32	4%
$100k to $500k	2	0%
$500k plus	3	0%
Unanswered	329	43%

Q21: Sales Tax Compliance Costs:
Programming and servicing cash registers and other POS systems to address sales-tax requirements

Category	*n*	*pct*
$0 (none)	188	25%
Under $50	10	1%
$50 to $100	10	1%
$100 to $500	71	9%
$500 to $1k	35	5%
$1k to $5k	68	9%
$5k to $10k	13	2%
$10k to $100k	30	4%
$100k to $500k	4	1%
$500k plus	4	1%
Unanswered	329	43%

Q22: Sales Tax Compliance Costs:
Returns preparation, making remittances, refund and credit claims, and research relating to sales tax?

Category	*n*	*pct*
$0 (none)	41	5%
Under $50	6	1%
$50 to $100	2	0%
$100 to $500	74	10%
$500 to $1k	58	8%

$1k to $5k	147	19%
$5k to $10k	41	5%
$10k to $100k	52	7%
$100k to $500k	9	1%
$500k plus	3	0%
Unanswered	329	43%

Q23: Sales Tax Compliance Costs:
Dealing with sales tax audits and appeals

Category	*n*	*pct*
$0 (none)	228	30%
Under $50	1	0%
$50 to $100	4	1%
$100 to $500	29	4%
$500 to $1k	19	2%
$1k to $5k	66	9%
$5k to $10k	27	4%
$10k to $100k	43	6%
$100k to $500k	10	1%
$500k plus	6	1%
Unanswered	329	43%

Q24: Sales Tax Compliance Costs:
Other costs not covered above (for example, costs related to data storage, sales tax registration, etc.)

Category	*n*	*pct*
$0 (none)	208	27%
Under $50	8	1%
$50 to $100	14	2%
$100 to $500	54	7%
$500 to $1k	41	5%
$1k to $5k	55	7%
$5k to $10k	20	3%
$10k to $100k	27	4%
$100k to $500k	4	1%
$500k plus	2	0%
Unanswered	329	43%

Q25: Sales Tax Compliance Costs:
If you are unable to break down your costs into the above categories, what is your best estimate of the total additional annual costs incurred because of the retail sales tax? (If you have provided answers to Questions 17 through 24, please ignore this question.)

Total annual sales tax compliance cost in 2003	*n*	*pct*
$0 (none)	91	12%
Under $50	7	1%
$50 to $100	2	0%
$100 to $500	60	8%
$500 to $1k	63	8%
$1k to $5k	216	28%
$5k to $10k	100	13%
$10k to $100k	166	22%
$100k to $500k	34	4%
$500k plus	18	2%
Unanswered	5	1%

Q26: Sales Tax Compliance Costs:
Of your sales tax compliance costs reported above in either Questions 17 through 24 or Question 25, about how much in total was for payments to outside service providers (lawyers, accountants, programmers, etc.)?

Portion of total sales tax compliance cost paid to outside service providers	*n*	*pct*
$0 (none)	399	52%
Under $50	0	0%
$50 to $100	2	0%
$100 to $500	48	6%
$500 to $1k	57	7%
$1k to $5k	142	19%
$5k to $10k	29	4%
$10k to $100k	49	6%
$100k to $500k	13	2%
$500k plus	5	1%
Unanswered	18	2%

Q27: How many state and local tax returns did you file in 2003?

Number of state sales tax returns	*n*	*pct*
Zero	21	3%
One	109	14%
Two to Five	89	12%
Six to Nine	4	1%
Ten to Twenty	335	44%
21 to 50	69	9%
51 to 100	20	3%
101 and up	46	6%
Unanswered	69	9%

Number of local sales tax returns	*n*	*pct*
Zero	240	32%
One	42	6%
Two to Five	42	6%
Six to Nine	5	1%
Ten to Twenty	94	12%
21 to 50	29	4%
51 to 100	13	2%
101 and up	34	4%
Unanswered	263	35%

Q28: To how many different jurisdictions did you submit sales tax returns in 2003?

Number of state jurisdictions	*n*	*pct*
Zero	23	3%
One	518	68%
Two to Five	80	11%
Six to Nine	11	1%
Ten to Twenty	14	2%
21 to 50	35	5%
51 to 100	0	0%
101 and up	1	0%
Unanswered	80	11%

Number of local jurisdictions	*n*	*pct*
Zero	206	27%
One	163	21%
Two to Five	58	8%
Six to Nine	14	2%
Ten to Twenty	18	2%
21 to 50	13	2%
51 to 100	11	1%

101 and up	14	2%
Unanswered	265	35%

Q29: Approximately how much did you remit in sales tax in 2003?

Amount of sales tax remitted	*n*	*pct*
$0 (none)	62	8%
Under $50	0	0%
$50 to $100	0	0%
$100 to $500	2	0%
$500 to $1k	4	1%
$1k to $5k	37	5%
$5k to $10k	26	3%
$10k to $100k	218	29%
$100k to $500k	166	22%
$500k plus	232	30%
Unanswered	15	2%

Q30: How much retail sales tax did you pay in 2003 that came out of your pocket because the customer defaulted and the state or local government would not provide a bad debt credit to you?

Unrecovered sales tax	*n*	*pct*
$0 (none)	498	65%
Under $50	9	1%
$50 to $100	6	1%
$100 to $500	64	8%
$500 to $1k	23	3%
$1k to $5k	66	9%
$5k to $10k	24	3%
$10k to $100k	30	4%
$100k to $500k	6	1%
$500k plus	8	1%
Unanswered	28	4%

Q31: Approximately how much of the sales tax you collected in 2003 were you allowed to retain as a discount for timely payment?

Vendor discount	*n*	*pct*
$0 (none)	329	43%
Under $50	17	2%

$50 to $100	10	1%
$100 to $500	72	9%
$500 to $1k	61	8%
$1k to $5k	109	14%
$5k to $10k	33	4%
$10k to $100k	58	8%
$100k to $500k	19	2%
$500k plus	16	2%
Unanswered	38	5%

Q32: What percent of your sales tax collections are received prior to remittance to the respective tax authorities?

Received prior to remittance	*n*	*pct*
NA: 0%	43	6%
Under 10%	18	2%
10% to 20%	7	1%
20% to 30%	15	2%
30% to 40%	9	1%
40% to 50%	33	4%
50% to 60%	14	2%
60% to 70%	14	2%
70% to 80%	44	6%
80% to 90%	46	6%
90% to 100%	424	56%
Unanswered	95	12%

Q32a: Of this amount, what is the average number of days between collection and remittance?

Days	*n*	*pct*
Under 5	59	8%
5 to 9	19	2%
10 to 14	64	8%
15 to 19	120	16%
20 to 29	135	18%
30 or more	216	28%
Unanswered	149	20%

Q33: What percent of your sales tax collections are received after remittance to the respective tax authorities?		
Received after remittance	*n*	*pct*
NA: 0%	0	0%
Under 10%	475	62%
10% to 20%	34	4%
20% to 30%	26	3%
30% to 40%	21	3%
40% to 50%	32	4%
50% to 60%	10	1%
60% to 70%	11	1%
70% to 80%	7	1%
80% to 90%	5	1%
90% to 100%	8	1%
Unanswered	133	17%

Q33a: Of this amount, what is the average number of days between remittance and collection?		
Days	*n*	*pct*
Under 5	185	24%
5 to 9	15	2%
10 to 14	22	3%
15 to 19	23	3%
20 to 29	36	5%
30 or more	143	19%
Unanswered	338	44%

Q34: Of tax-exempt sales in 2003, what percent was related to each of the following reasons?		
Nontaxable goods and services	*n*	*pct*
Exactly 0%	225	30%
0% to 10%	70	9%
10% to 20%	29	4%
20% to 30%	24	3%
30% to 40%	17	2%
40% to 50%	29	4%
50% to 60%	15	2%
60% to 70%	14	2%
70% to 80%	20	3%
80% to 90%	33	4%
90% to 100%	135	18%
Unanswered	151	20%

Resale certificates	*n*	*pct*
Exactly 0%	224	29%
0% to 10%	116	15%
10% to 20%	40	5%
20% to 30%	30	4%
30% to 40%	20	3%
40% to 50%	28	4%
50% to 60%	16	2%
60% to 70%	24	3%
70% to 80%	30	4%
80% to 90%	30	4%
90% to 100%	53	7%
Unanswered	151	20%

Out-of-state sales to non-nexus states	*n*	*pct*
Exactly 0%	422	55%
0% to 10%	98	13%
10% to 20%	24	3%
20% to 30%	13	2%
30% to 40%	6	1%
40% to 50%	6	1%
50% to 60%	5	1%
60% to 70%	5	1%
70% to 80%	5	1%
80% to 90%	10	1%
90% to 100%	17	2%
Unanswered	151	20%

Sales to exempt organizations	*n*	*pct*
Exactly 0%	228	30%
0% to 10%	207	27%
10% to 20%	45	6%
20% to 30%	22	3%
30% to 40%	16	2%
40% to 50%	20	3%
50% to 60%	6	1%
60% to 70%	2	0%
70% to 80%	14	2%
80% to 90%	10	1%
90% to 100%	41	5%
Unanswered	151	20%

Other exempt sales	*n*	*pct*
Exactly 0%	460	60%
0% to 10%	68	9%
10% to 20%	16	2%
20% to 30%	6	1%
30% to 40%	9	1%

40% to 50%	8	1%
50% to 60%	7	1%
60% to 70%	10	1%
70% to 80%	3	0%
80% to 90%	4	1%
90% to 100%	20	3%
Unanswered	151	20%

Q35: Approximately what percent of your sales tax documentation (including documentation for exempt sales) was stored as follows?

Electronically	*n*	*pct*
0% to 10%	408	54%
10% to 20%	7	1%
20% to 30%	7	1%
30% to 40%	6	1%
40% to 50%	62	8%
50% to 60%	4	1%
60% to 70%	5	1%
70% to 80%	23	3%
80% to 90%	27	4%
90% to 100%	119	16%
Unanswered	94	12%

In paper files	*n*	*pct*
0% to 10%	150	20%
10% to 20%	20	3%
20% to 30%	16	2%
30% to 40%	2	0%
40% to 50%	61	8%
50% to 60%	5	1%
60% to 70%	6	1%
70% to 80%	11	1%
80% to 90%	11	1%
90% to 100%	386	51%
Unanswered	94	12%

In other ways	*n*	*pct*
0% to 10%	656	86%
10% to 20%	1	0%
20% to 30%	0	0%
30% to 40%	1	0%
40% to 50%	2	0%
50% to 60%	0	0%
60% to 70%	0	0%

70% to 80%	0	0%
80% to 90%	0	0%
90% to 100%	8	1%
Unanswered	94	12%

Q36: How many sales tax audits were either started or ongoing in 2003? (Exclude use tax audits)

Number of audits in 2003 (if any)	*n*	*pct*
Zero	496	65%
One	110	14%
Two	10	1%
Three	7	1%
Four	4	1%
Five or more	27	4%
Unanswered	108	14%

Q37: How many years do your sales tax audits typically cover?

Number of years	*n*	*pct*
One	42	6%
Two	32	4%
Three	230	30%
Four	54	7%
Five	43	6%
Six or more	28	4%
Unanswered	333	44%

Q38: Do you currently have any ongoing appeals of a sales tax audit finding?

Answer	*n*	*pct*
Yes	39	5%
No	646	85%
Unanswered	77	10%

Q38a: If yes, how many appeals?

Number of ongoing appeals	*n*	*pct*
One	20	3%
Two	8	1%
Three	3	0%

Four	2	0%
Five	0	0%
Six or more	4	1%
Unanswered	725	95%

Q39: Describe the nature of your primary type of business

Category	*n*	*pct*
Answered	709	93%
Unanswered	53	7%

Q40: How long have you been in business in the United States (check one)?

Time	*n*	*pct*
Less than three years	16	2%
Three years or more	705	93%
Unanswered	41	5%

Part II – Fully Edited But Not Imputed Survey Data by Question
(JCCS Full Sample of 796 Observations)

Q1: What do you consider the greatest cost in collecting and remitting sales tax?

Category	*n*	*pct*
Training personnel	38	5%
Programming cash register/POS systems	51	6%
Preparing tax forms including research	335	42%
Remitting sales tax	67	8%
Handling audits	46	6%
Keeping track of local taxes	46	6%
Documenting exempt sales	88	11%
Other	48	6%
Unanswered	77	10%

Q2: About how many retail sales transactions (invoices) did you have per day in 2003?

Average number of daily retail sales transactions	*n*	*pct*
0 to 10	78	10%
10 to 100	277	35%
100 to 1000	227	29%
1000 to 10,000	75	9%
10,000 plus	52	7%
Unanswered	87	11%

Q3: How many different products (e.g., SKUs) did you sell at retail as of year-end 2003?

Category	*n*	*pct*
Less than 1,000	298	37%
1,000 to 5,000	161	20%
5,000 to 10,000	90	11%
10,000 to 25,000	85	11%
25,000 to 50,000	51	6%
50,000 to 100,000	38	5%
100,000 or more	36	5%
Unanswered	37	5%

Q4: What percent of your retail sales dollars in 2003 were through the following channels?		
Retail store sales	*n*	*pct*
0% to 10%	24	3%
11% to 20%	-	0%
21% to 30%	6	1%
31% to 40%	3	0%
41% to 50%	2	0%
51% to 60%	2	0%
61% to 70%	3	0%
71% to 80%	14	2%
81% to 90%	28	4%
91% to 100%	676	85%
Unanswered	38	5%
Catalogue sales	*n*	*pct*
0% to 10%	713	90%
11% to 20%	9	1%
21% to 30%	6	1%
31% to 40%	3	0%
41% to 50%	3	0%
51% to 60%	2	0%
61% to 70%	4	1%
71% to 80%	3	0%
81% to 90%	5	1%
91% to 100%	10	1%
Unanswered	38	5%
Internet sales	*n*	*pct*
0% to 10%	734	92%
11% to 20%	10	1%
21% to 30%	8	1%
31% to 40%	-	0%
41% to 50%	-	0%
51% to 60%	-	0%
61% to 70%	1	0%
71% to 80%	-	0%
81% to 90%	-	0%
91% to 100%	5	1%
Unanswered	38	5%

Q5: Please provide the following information for your U.S. retail activities in 2003:		
Gross sales before returns and allowances	*n*	*pct*
Under $150k	39	5%

$150k to $200k	6	1%
$200k to $500k	60	8%
$500k to $1M	49	6%
$1M to $2.5M	91	11%
$2.5M to $5M	64	8%
$5M to $10M	62	8%
$10M to $25M	81	10%
$25M to $50M	72	9%
$50M to $100M	60	8%
$100M plus	84	11%
Unanswered	128	16%

Taxable sales	*n*	*pct*
Under $150k	109	14%
$150k to $200k	20	3%
$200k to $500k	69	9%
$500k to $1M	43	5%
$1M to $2.5M	82	10%
$2.5M to $5M	74	9%
$5M to $10M	58	7%
$10M to $25M	73	9%
$25M to $50M	50	6%
$50M to $100M	38	5%
$100M plus	55	7%
Unanswered	125	16%

Q6: How much were your remote sales (catalogue or Internet), if any, in 2003?

Shipments to all US locations	*n*	*pct*
NA: $0 Entered	467	59%
Under $150k	43	5%
$150k to $200k	3	0%
$200k to $500k	5	1%
$500k to $1M	6	1%
$1M to $2.5M	9	1%
$2.5M to $5M	7	1%
$5M to $10M	4	1%
$10M to $25M	14	2%
$25M to $50M	6	1%
$50M to $100M	2	0%
$100M plus	15	2%
Unanswered	215	27%

	n	*pct*

Shipments on which you collect and remit sales tax		
NA: $0 Entered	463	58%
Under $150k	40	5%
$150k to $200k	1	0%
$200k to $500k	4	1%
$500k to $1M	5	1%
$1M to $2.5M	7	1%
$2.5M to $5M	5	1%
$5M to $10M	7	1%
$10M to $25M	11	1%
$25M to $50M	1	0%
$50M to $100M	1	0%
$100M plus	12	2%
Unanswered	239	30%

Q7: How many states (including the District of Columbia) did you ship to in 2003?

Number of states shipped to	*n*	*pct*
Zero	273	34%
One	145	18%
Two	47	6%
Three to Five	85	11%
Six t Nine	23	3%
Ten to Twenty	43	5%
21 to 40	15	2%
41 and up	47	6%
Unanswered	118	15%

Number of states shipped to for which you collect and remit sales tax	*n*	*pct*
Zero	330	41%
One	179	22%
Two	29	4%
Three to Five	43	5%
Six t Nine	13	2%
Ten to Twenty	5	1%
21 to 40	16	2%
41 and up	16	2%
Unanswered	165	21%

Q8: Please indicate below the number of retail stores you had in each state (including the District of Columbia), if any, as of June 30, 2003:

State	*n*	*pct*
AL:	650	2%
AK:	158	0%
AZ:	841	2%
AR:	435	1%
CA:	4,144	10%
CO:	620	2%
CT:	468	1%
DE:	126	0%
DC:	41	0%
FL:	3,039	7%
GA:	1,190	3%
HI:	175	0%
ID:	234	1%
IL:	2,245	6%
IN:	1,054	3%
IA:	474	1%
KS:	390	1%
KY:	599	1%
LA:	677	2%
ME:	149	0%
MD:	621	2%
MA:	907	2%
MI:	1,457	4%
MN:	672	2%
MS:	511	1%
MO:	856	2%
MT:	183	0%
NE:	267	1%
NV:	390	1%
NH:	257	1%
NJ:	1,082	3%
NM:	279	1%
NY:	1,880	5%
NC:	1,081	3%
ND:	116	0%
OH:	1,849	5%
OK:	450	1%
OR:	455	1%
PA:	1,443	4%
RI:	143	0%
SC:	521	1%
SD:	107	0%
TN:	883	2%
TX:	3,367	8%
UT:	393	1%
VT:	75	0%
VA:	894	2%

WA:	856	2%
WV:	180	0%
WI:	835	2%
WY:	81	0%

Q9: How many employees did you have at the end of 2003?

Employees in tax department (fraction if less than 1)	*n*	*pct*
Zero	133	17%
Fraction under 1	128	16%
1 to 2.9	250	31%
3 to 5	34	4%
6 to 9	9	1%
10 to 20	19	2%
21 to 40	10	1%
41 or higher	17	2%
Unanswered	196	25%

Employees in accounting department (fraction if less than 1)	*n*	*pct*
Zero	50	6%
Fraction under 1	62	8%
1 to 2.9	248	31%
3 to 5	167	21%
6 to 9	66	8%
10 to 20	49	6%
21 to 40	18	2%
41 or higher	29	4%
Unanswered	107	13%

Employees in customer service department (fraction if less than 1)	*n*	*pct*
Zero	76	10%
Fraction under 1	20	3%
1 to 2.9	185	23%
3 to 5	121	15%
6 to 9	61	8%
10 to 20	71	9%
21 to 40	39	5%
41 or higher	58	7%
Unanswered	165	21%

Cashiers		
(fraction if less than 1)	*n*	*pct*
Zero	86	11%
Fraction under 1	23	3%
1 to 2.9	228	29%
3 to 5	104	13%
6 to 9	37	5%
10 to 20	49	6%
21 to 40	27	3%
41 or higher	88	11%
Unanswered	154	19%

Other employees		
(fraction if less than 1)	*n*	*pct*
Zero	44	6%
Fraction under 1	5	1%
1 to 2.9	76	10%
3 to 5	52	7%
6 to 9	39	5%
10 to 20	87	11%
21 to 40	91	11%
41 or higher	251	32%
Unanswered	151	19%

Q10: About what percent of the gross sales of your retail business were:

Returned or exchanged	*n*	*pct*
NA: 0%	163	20%
Under 10%	466	59%
10% to 20%	20	3%
20% to 30%	4	1%
30% to 40%	2	0%
40% to 50%	-	0%
50% to 60%	-	0%
60% to 70%	-	0%
70% to 80%	1	0%
80% to 90%	-	0%
90% to 100%	2	0%
Unanswered	138	17%

Written off as a bad debt	*n*	*pct*
NA: 0%	233	29%
Under 10%	403	51%
10% to 20%	5	1%
20% to 30%	2	0%
30% to 40%	-	0%

40% to 50%	-	0%
50% to 60%	-	0%
60% to 70%	-	0%
70% to 80%	-	0%
80% to 90%	-	0%
90% to 100%	-	0%
Unanswered	153	19%

Q11: Approximately what percent of your total sales dollars were paid in the following ways in 2003?

Cash	*n*	*pct*
0% to 10%	324	41%
10% to 20%	124	16%
20% to 30%	89	11%
30% to 40%	52	7%
40% to 50%	42	5%
50% to 60%	20	3%
60% to 70%	19	2%
70% to 80%	16	2%
80% to 90%	9	1%
90% to 100%	8	1%
Unanswered	93	12%

Checks	*n*	*pct*
0% to 10%	167	21%
10% to 20%	106	13%
20% to 30%	91	11%
30% to 40%	65	8%
40% to 50%	69	9%
50% to 60%	33	4%
60% to 70%	25	3%
70% to 80%	55	7%
80% to 90%	33	4%
90% to 100%	59	7%
Unanswered	93	12%

Debit cards	*n*	*pct*
0% to 10%	594	75%
10% to 20%	68	9%
20% to 30%	29	4%
30% to 40%	7	1%
40% to 50%	1	0%
50% to 60%	1	0%
60% to 70%	2	0%
70% to 80%	-	0%

80% to 90%	1	0%
90% to 100%	-	0%
Unanswered	93	12%
In-House credit cards	*n*	*pct*
0% to 10%	650	82%
10% to 20%	21	3%
20% to 30%	9	1%
30% to 40%	7	1%
40% to 50%	6	1%
50% to 60%	4	1%
60% to 70%	1	0%
70% to 80%	3	0%
80% to 90%	1	0%
90% to 100%	1	0%
Unanswered	93	12%
Other credit cards	*n*	*pct*
0% to 10%	289	36%
10% to 20%	126	16%
20% to 30%	97	12%
30% to 40%	60	8%
40% to 50%	67	8%
50% to 60%	24	3%
60% to 70%	19	2%
70% to 80%	13	2%
80% to 90%	5	1%
90% to 100%	3	0%
Unanswered	93	12%
Other	*n*	*pct*
0% to 10%	596	75%
10% to 20%	16	2%
20% to 30%	13	2%
30% to 40%	5	1%
40% to 50%	9	1%
50% to 60%	14	2%
60% to 70%	20	3%
70% to 80%	15	2%
80% to 90%	9	1%
90% to 100%	6	1%
Unanswered	93	12%

Q12: For each of the following types of payment, indicate the average percentage fee you paid to the credit card company or other financial institution in 2003:

Fee for debit cards	*n*	*pct*
0% to .5%	118	15%
.5% to 1%	22	3%
1% to 1.5%	50	6%
1.5% to 2%	67	8%
2% to 3%	120	15%
3% to 4%	50	6%
4% to 5%	7	1%
5% to 10%	7	1%
10% to 20%	-	0%
20% to 100%	-	0%
Unanswered	355	45%

Fee for in-house credits cards	*n*	*pct*
0% to .5%	190	24%
.5% to 1%	1	0%
1% to 1.5%	4	1%
1.5% to 2%	17	2%
2% to 3%	39	5%
3% to 4%	11	1%
4% to 5%	3	0%
5% to 10%	1	0%
10% to 20%	-	0%
20% to 100%	-	0%
Unanswered	530	67%

Fee for other credit cards	*n*	*pct*
0% to .5%	62	8%
.5% to 1%	3	0%
1% to 1.5%	20	3%
1.5% to 2%	121	15%
2% to 3%	260	33%
3% to 4%	123	15%
4% to 5%	15	2%
5% to 10%	6	1%
10% to 20%	-	0%
20% to 100%	-	0%
Unanswered	186	23%

Q13: How many cash registers (including POS terminals and cash box/calculators) did you use in 2003?

Number of cash registers	*n*	*pct*
Zero	64	8%
One	237	30%
Two	128	16%
Three to Five	132	17%
Six to Nine	43	5%
Ten to Twenty	50	6%
21 to 50	20	3%
50 and up	80	10%
Unanswered	42	5%

Q14: About what percent of all cash registers used by your employees were of the following types in 2003?

Manual (including cash box and calculator):

Percentage Range	*n*	*pct*
0% to 10%	416	52%
10% to 20%	4	1%
20% to 30%	2	0%
30% to 40%	8	1%
40% to 50%	10	1%
50% to 60%	2	0%
60% to 70%	-	0%
70% to 80%	-	0%
80% to 90%	1	0%
90% to 100%	259	33%
Unanswered	94	12%

Semi-manual (without electronic data files):

Percentage Range	*n*	*pct*
0% to 10%	572	72%
10% to 20%	5	1%
20% to 30%	-	0%
30% to 40%	3	0%
40% to 50%	8	1%
50% to 60%	-	0%
60% to 70%	1	0%
70% to 80%	2	0%
80% to 90%	1	0%
90% to 100%	110	14%
Unanswered	94	12%

Automatic registers with electronic data files:		
Percentage Range	*n*	*pct*
0% to 10%	381	48%
10% to 20%	1	0%
20% to 30%	-	0%
30% to 40%	1	0%
40% to 50%	6	1%
50% to 60%	-	0%
60% to 70%	7	1%
70% to 80%	7	1%
80% to 90%	4	1%
90% to 100%	295	37%
Unanswered	94	12%

Q15: What was the approximate cost for a new cash register of each type that you used in your retail business in 2003?

Manual	*n*	*pct*
NA	-	0%
$0 to $5	1	0%
$5 to $20	8	1%
$20 to $50	16	2%
$50 to $100	39	5%
$100 to $200	42	5%
$200 to $500	52	7%
$500 to $1000	20	3%
$1000 to $2000	6	1%
$2000 to $5000	3	0%
$5000 and up	1	0%
Unanswered	608	76%

Semi-manual	*n*	*pct*
NA	-	0%
$0 to $5	-	0%
$5 to $20	-	0%
$20 to $50	1	0%
$50 to $100	2	0%
$100 to $200	13	2%
$200 to $500	29	4%
$500 to $1000	32	4%
$1000 to $2000	9	1%
$2000 to $5000	13	2%
$5000 and up	5	1%
Unanswered	692	87%

Automatic	*n*	*pct*
NA	-	0%
$0 to $5	1	0%
$5 to $20	-	0%
$20 to $50	2	0%
$50 to $100	1	0%
$100 to $200	4	1%
$200 to $500	21	3%
$500 to $1000	41	5%
$1000 to $2000	63	8%
$2000 to $5000	77	10%
$5000 to $10000	39	5%
$10000 and up	31	4%
Unanswered	516	65%

Q16: Would you have purchased less expensive cash registers if there were no sales tax?

Answer	*n*	*pct*
Yes	71	9%
No	571	72%
Unanswered	154	19%

Q16a: If yes, approximately how much more do you pay for a typical cash register than you would have paid in the absence of any sales tax?

Category	*n*	*pct*
NA	725	91%
$0 to $5	-	0%
$5 to $20	-	0%
$20 to $50	1	0%
$50 to $100	3	0%
$100 to $200	6	1%
$200 to $500	15	2%
$500 and up	31	4%
Unanswered	15	2%

Q17: Sales Tax Compliance Costs: Training Personnel on sales tax

Category	*n*	*pct*
$0 (none)	121	15%
Under $50	3	0%

$50 to $100	11	1%
$100 to $500	97	12%
$500 to $1k	42	5%
$1k to $5k	99	12%
$5k to $10k	26	3%
$10k to $100k	37	5%
$100k to $500k	1	0%
$500k plus	4	1%
Unanswered	355	45%

Q18: Sales Tax Compliance Costs:
Documenting tax-exempt sales

Category	*n*	*pct*
$0 (none)	74	9%
Under $50	14	2%
$50 to $100	9	1%
$100 to $500	108	14%
$500 to $1k	41	5%
$1k to $5k	108	14%
$5k to $10k	33	4%
$10k to $100k	45	6%
$100k to $500k	3	0%
$500k plus	6	1%
Unanswered	355	45%

Q19: Sales Tax Compliance Costs:
Customer service relating to sales tax issues
other than documenting exempt sales

Category	*n*	*pct*
$0 (none)	205	26%
Under $50	7	1%
$50 to $100	11	1%
$100 to $500	70	9%
$500 to $1k	30	4%
$1k to $5k	75	9%
$5k to $10k	19	2%
$10k to $100k	18	2%
$100k to $500k	3	0%
$500k plus	3	0%
Unanswered	355	45%

Q20: Sales Tax Compliance Costs:
Sales tax-related software acquisitions and license fees

Category	*n*	*pct*
$0 (none)	224	28%
Under $50	4	1%
$50 to $100	7	1%
$100 to $500	63	8%
$500 to $1k	34	4%
$1k to $5k	65	8%
$5k to $10k	7	1%
$10k to $100k	32	4%
$100k to $500k	2	0%
$500k plus	3	0%
Unanswered	355	45%

Q21: Sales Tax Compliance Costs:
Programming and servicing cash registers and other POS systems to address sales-tax requirements

Category	*n*	*pct*
$0 (none)	194	24%
Under $50	10	1%
$50 to $100	11	1%
$100 to $500	71	9%
$500 to $1k	35	4%
$1k to $5k	68	9%
$5k to $10k	13	2%
$10k to $100k	31	4%
$100k to $500k	4	1%
$500k plus	4	1%
Unanswered	355	45%

Q22: Sales Tax Compliance Costs:
Returns preparation, making remittances, refund and credit claims, and research relating to sales tax?

Category	*n*	*pct*
$0 (none)	42	5%
Under $50	7	1%
$50 to $100	2	0%
$100 to $500	75	9%
$500 to $1k	58	7%

$1k to $5k	151	19%
$5k to $10k	41	5%
$10k to $100k	53	7%
$100k to $500k	9	1%
$500k plus	3	0%
Unanswered	355	45%

Q23: Sales Tax Compliance Costs:
Dealing with sales tax audits and appeals

Category	*n*	*pct*
$0 (none)	233	29%
Under $50	1	0%
$50 to $100	4	1%
$100 to $500	29	4%
$500 to $1k	20	3%
$1k to $5k	66	8%
$5k to $10k	29	4%
$10k to $100k	43	5%
$100k to $500k	10	1%
$500k plus	6	1%
Unanswered	355	45%

Q24: Sales Tax Compliance Costs:
Other costs not covered above (for example, costs related to data storage, sales tax registration, etc.)

Category	*n*	*pct*
$0 (none)	211	27%
Under $50	8	1%
$50 to $100	14	2%
$100 to $500	54	7%
$500 to $1k	41	5%
$1k to $5k	60	8%
$5k to $10k	20	3%
$10k to $100k	27	3%
$100k to $500k	4	1%
$500k plus	2	0%
Unanswered	355	45%

Q25: Sales Tax Compliance Costs:
If you are unable to break down your costs into the above categories, what is your best estimate of the total additional annual costs incurred because of the retail sales tax? (If you have provided answers to Questions 17 through 24, please ignore this question.)

Total annual sales tax compliance cost in 2003	*n*	*pct*
$0 (none)	31	4%
Under $50	5	1%
$50 to $100	2	0%
$100 to $500	59	7%
$500 to $1k	57	7%
$1k to $5k	202	25%
$5k to $10k	98	12%
$10k to $100k	155	19%
$100k to $500k	32	4%
$500k plus	18	2%
Unanswered	137	17%

Q26: Sales Tax Compliance Costs:
Of your sales tax compliance costs reported above in either Questions 17 through 24 or Question 25, about how much in total was for payments to outside service providers (lawyers, accountants, programmers, etc.)?

Portion of total sales tax compliance cost paid to outside service providers	*n*	*pct*
$0 (none)	289	36%
Under $50	1	0%
$50 to $100	2	0%
$100 to $500	37	5%
$500 to $1k	44	6%
$1k to $5k	122	15%
$5k to $10k	24	3%
$10k to $100k	47	6%
$100k to $500k	9	1%
$500k plus	4	1%
Unanswered	217	27%

Q27: How many state and local tax returns did you file in 2003?

Number of state sales tax returns	*n*	*pct*
Zero	22	3%
One	116	15%
Two to Five	93	12%
Six to Nine	6	1%
Ten to Twenty	343	43%
21 to 50	71	9%
51 to 100	20	3%
101 and up	46	6%
Unanswered	79	10%

Number of local sales tax returns	*n*	*pct*
Zero	250	31%
One	45	6%
Two to Five	42	5%
Six to Nine	5	1%
Ten to Twenty	98	12%
21 to 50	29	4%
51 to 100	13	2%
101 and up	34	4%
Unanswered	280	35%

Q28: To how many different jurisdictions did you submit sales tax returns in 2003?

Number of state jurisdictions	*n*	*pct*
Zero	24	3%
One	535	67%
Two to Five	84	11%
Six to Nine	11	1%
Ten to Twenty	14	2%
21 to 50	35	4%
51 to 100	-	0%
101 and up	1	0%
Unanswered	92	12%

Number of local jurisdictions	*n*	*pct*
Zero	215	27%
One	169	21%
Two to Five	59	7%
Six to Nine	14	2%
Ten to Twenty	19	2%
21 to 50	13	2%
51 to 100	11	1%

101 and up	14	2%
Unanswered	282	35%

Q29: Approximately how much did you remit in sales tax in 2003?

Amount of sales tax remitted	*n*	*pct*
$0 (none)	18	2%
Under $50	1	0%
$50 to $100	-	0%
$100 to $500	2	0%
$500 to $1k	6	1%
$1k to $5k	31	4%
$5k to $10k	21	3%
$10k to $100k	191	24%
$100k to $500k	155	19%
$500k plus	220	28%
Unanswered	151	19%

Q30: How much retail sales tax did you pay in 2003 that came out of your pocket because the customer defaulted and the state or local government would not provide a bad debt credit to you?

Unrecovered sales tax	*n*	*pct*
$0 (none)	389	49%
Under $50	10	1%
$50 to $100	5	1%
$100 to $500	53	7%
$500 to $1k	21	3%
$1k to $5k	63	8%
$5k to $10k	20	3%
$10k to $100k	26	3%
$100k to $500k	6	1%
$500k plus	8	1%
Unanswered	195	24%

Q31: Approximately how much of the sales tax you collected in 2003 were you allowed to retain as a discount for timely payment?

Vendor discount	*n*	*pct*
$0 (none)	268	34%
Under $50	17	2%

$50 to $100	11	1%
$100 to $500	67	8%
$500 to $1k	53	7%
$1k to $5k	94	12%
$5k to $10k	30	4%
$10k to $100k	54	7%
$100k to $500k	19	2%
$500k plus	15	2%
Unanswered	168	21%

Q32: What percent of your sales tax collections are received prior to remittance to the respective tax authorities?

Received prior to remittance	*n*	*pct*
NA: 0%	49	6%
Under 10%	19	2%
10% to 20%	7	1%
20% to 30%	15	2%
30% to 40%	9	1%
40% to 50%	34	4%
50% to 60%	14	2%
60% to 70%	14	2%
70% to 80%	46	6%
80% to 90%	47	6%
90% to 100%	435	55%
Unanswered	107	13%

Q32a: Of this amount, what is the average number of days between collection and remittance?

Days	*n*	*pct*
Under 5	64	8%
5 to 9	21	3%
10 to 14	66	8%
15 to 19	121	15%
20 to 29	138	17%
30 or more	222	28%
Unanswered	164	21%

Q33: What percent of your sales tax collections are received after remittance to the respective tax authorities?

Received after remittance	*n*	*pct*
NA: 0%	-	0%
Under 10%	491	62%
10% to 20%	35	4%
20% to 30%	27	3%
30% to 40%	21	3%
40% to 50%	33	4%
50% to 60%	10	1%
60% to 70%	11	1%
70% to 80%	7	1%
80% to 90%	5	1%
90% to 100%	11	1%
Unanswered	145	18%

Q33a: Of this amount, what is the average number of days between remittance and collection?

Days	*n*	*pct*
Under 5	192	24%
5 to 9	15	2%
10 to 14	22	3%
15 to 19	23	3%
20 to 29	36	5%
30 or more	150	19%
Unanswered	358	45%

Q34: Of tax-exempt sales in 2003, what percent was related to each of the following reasons?

Nontaxable goods and services	*n*	*pct*
Exactly 0%	234	29%
0% to 10%	70	9%
10% to 20%	30	4%
20% to 30%	24	3%
30% to 40%	17	2%
40% to 50%	30	4%
50% to 60%	15	2%
60% to 70%	14	2%
70% to 80%	21	3%
80% to 90%	33	4%

90% to 100%	141	18%
Unanswered	167	21%
Resale certificates	*n*	*pct*
Exactly 0%	235	30%
0% to 10%	116	15%
10% to 20%	40	5%
20% to 30%	31	4%
30% to 40%	21	3%
40% to 50%	28	4%
50% to 60%	16	2%
60% to 70%	24	3%
70% to 80%	30	4%
80% to 90%	30	4%
90% to 100%	58	7%
Unanswered	167	21%
Out-of-state sales to non-nexus states	*n*	*pct*
Exactly 0%	436	55%
0% to 10%	101	13%
10% to 20%	24	3%
20% to 30%	13	2%
30% to 40%	6	1%
40% to 50%	6	1%
50% to 60%	5	1%
60% to 70%	5	1%
70% to 80%	5	1%
80% to 90%	10	1%
90% to 100%	18	2%
Unanswered	167	21%
Sales to exempt organizations	*n*	*pct*
Exactly 0%	242	30%
0% to 10%	208	26%
10% to 20%	45	6%
20% to 30%	23	3%
30% to 40%	16	2%
40% to 50%	21	3%
50% to 60%	7	1%
60% to 70%	2	0%
70% to 80%	14	2%
80% to 90%	10	1%
90% to 100%	41	5%
Unanswered	167	21%
	n	*pct*

Other exempt sales		
Exactly 0%	475	60%
0% to 10%	68	9%
10% to 20%	16	2%
20% to 30%	6	1%
30% to 40%	9	1%
40% to 50%	8	1%
50% to 60%	8	1%
60% to 70%	10	1%
70% to 80%	4	1%
80% to 90%	4	1%
90% to 100%	21	3%
Unanswered	167	21%

Q35: Approximately what percent of your sales tax documentation (including documentation for exempt sales) was stored as follows?

Electronically	*n*	*pct*
0% to 10%	422	53%
10% to 20%	7	1%
20% to 30%	7	1%
30% to 40%	7	1%
40% to 50%	65	8%
50% to 60%	4	1%
60% to 70%	5	1%
70% to 80%	24	3%
80% to 90%	27	3%
90% to 100%	121	15%
Unanswered	107	13%

In paper files	*n*	*pct*
0% to 10%	152	19%
10% to 20%	20	3%
20% to 30%	17	2%
30% to 40%	2	0%
40% to 50%	64	8%
50% to 60%	5	1%
60% to 70%	7	1%
70% to 80%	11	1%
80% to 90%	11	1%
90% to 100%	400	50%
Unanswered	107	13%

In other ways	*n*	*pct*
0% to 10%	677	85%
10% to 20%	1	0%
20% to 30%	-	0%
30% to 40%	1	0%
40% to 50%	2	0%
50% to 60%	-	0%
60% to 70%	-	0%
70% to 80%	-	0%
80% to 90%	-	0%
90% to 100%	8	1%
Unanswered	107	13%

Q36: How many sales tax audits were either started or ongoing in 2003? (Exclude use tax audits)

Number of audits in 2003 (if any)	*n*	*pct*
Zero	516	65%
One	111	14%
Two	11	1%
Three	7	1%
Four	4	1%
Five or more	27	3%
Unanswered	120	15%

Q37: How many years do your sales tax audits typically cover?

Number of years	*n*	*pct*
One	42	5%
Two	32	4%
Three	235	30%
Four	55	7%
Five	43	5%
Six or more	28	4%
Unanswered	361	45%

Q38: Do you currently have any ongoing appeals of a sales tax audit finding?

Answer	*n*	*pct*
Yes	40	5%
No	668	84%
Unanswered	88	11%

Q38a: If yes, how many appeals?

Number of ongoing appeals	*n*	*pct*
One	21	3%
Two	8	1%
Three	3	0%
Four	2	0%
Five	-	0%
Six or more	4	1%
Unanswered	758	95%

Q39: Describe the nature of your primary type of business

Category	*n*	*pct*
Answered	738	93%
Unanswered	58	7%

Q40: How long have you been in business in the United States (check one)?

Time	*n*	*pct*
Less than three years	18	2%
Three years or more	730	92%
Unanswered	48	6%

III. CONSISTENCY CHECKS

This section summarizes the consistency checks performed on the survey data. Included here is a list of PwC-required consistency tests for all survey responses. Marginal information, by consistency check, for both failure counts and missingness counts is also presented. Note that the missingness is not a count of the number of missing survey items – it is simply a count of consistency checks with too much missing information to perform an imputation.

See the "Consistency of Data" subsection under Section I of Volume Two of this report for details on what was done as a result of edit failures.

Label	Consistency Check
Q02cc	Q2TRANS x 365 <= upper_bound (Q3PROD)
Q04cc	Q4RET+Q4CAT+Q4NET=100%[11]
Q05cc	Q5TAXSL <= Q5GROSS
Q06c1	Q6SHPTX<= Q6USSHP;
Q06c2	Q6USSHP<=Q5GROSS;
Q06c3	Q6SHPTX<=Q5TAXSL;
Q07c1	Q7USSHP<=Q7SHPTX;
Q07c2	If Q6USSHP>0 then Q7USSHP>0; If Q6SHPTX>0 then Q7SHPTX>0;
Q11cc	Q11CASH+Q11CHKS+Q11DEBIT+Q11CARD+Q11OTHCR+Q11OTH=100%
Q12c1	Q12DCARD<=5%; Q12CARD<=5%; Q12OTHCR<=5%
Q12c2	If Q12DCARD > 0 then Q11DEBIT > 0; If Q12CARD > 0 then Q11CARD > 0; If Q12OTHCR > 0 then Q11OTHCR > 0;
Q13cc	Q9CASHR >= Q13REG

[11] These consistency tests, where parts are expected to sum to 100%, are allowed to deviate up to 1%. So the sum must be between 99 and 101, non-inclusive.

Label	Consistency Check
Q14cc	Q14MAN+Q14SEMI+Q14AUTO=100%
Q15c1	Q15MAN < Q15SEMI < Q15AUTO;
Q15c2	If Q15MAN>0 then Q14MAN>0; If Q15SEMI>0 then Q14SEMI>0; If Q15AUTO>0 then Q14AUTO>0;
Q16cc	If Q13REG = 0 then Q16! = Yes
Q25cc	If Q25 and sum (Q17…Q24) both exist, then sum (Q17…Q24) is within $1 of Q25[12]
Q26cc	Q26PAY <= Q25EST or Q26PAY <= sum (Q17 … Q24)
Q27cc	If Q5TAXSL=0 then Q27STAX=0; If Q5TAXSL=0 then Q27LTAX=0;
Q28c1	Q27STAX>=Q28SJUR; Q27LTAX>=Q28LJUR;
Q28c2	If Q5TAXSL=0 then Q28SJUR=0; If Q5TAXSL=0 then Q28LJUR=0;
Q29cc	Q29REMIT <= 10% x Q5TAXSL
Q30cc	Q30TAX <= Q29REMIT
Q31cc	Q31VEN <= 5% x Q29REMIT
Q34cc	Q34NGOOD+Q34CERT+Q34OUT+Q34EXMP+Q34OTH=100%

[12] This check was not on the PWC list, but was performed as part of the imputation. When the check failed badly (off by more than 20%), all items Q17…Q25 were discarded and Q25 subsequently imputed. If the comparisons were "within range," then the breakdown (Q17 … Q24) was used and Q25 was replaced with the sum (Q17…Q24).

Summary Results of the Consistency Checks

Consistency Check Label	Total Consistency Failure per Consistency Check	Missingless per Consistency Check
Q02cc	35	129
Q04cc	18	20
Q05cc	8	158
Q11cc	24	57
Q12c1	30	118
Q12c2	0	135
Q13cc	188	168
Q14cc	54	69
Q16cc	0	161
Q25cc	44	118
Q26cc	9	218
Q29cc	67	180
Q30cc	1	215
Q31cc	17	185
Q34cc	41	123
Q35cc	16	100

IV. STANDARD ERROR ANALYSIS

1. Introduction

In what follows we describe the steps taken to calculate sampling variances. Four tables in Volume I of this report were chosen by the JCCS Steering Committee for this analysis and each cell had its sampling error estimated.[13] The basic replicate sampling error calculations were done first. An adjustment was then made for the fact that missing data had to be imputed. A short example of the use of the sampling errors that were obtained is given and then a summary concludes the text. References end this section.

2. Replicate Sampling Error

The method of random groups (Wolter 1985) was used to estimate variance due to sampling error. The population of 724 companies (cases below $150k were removed, so the number of cases has been reduced from 762) was randomly divided into 11 replicates that each represent the population equally well. The random division was stratified by the original sampling strata to ensure equal representation. One certainty case is represented in all replicates. The weights were then appropriately adjusted to ensure each replicate acted as a "miniature sample." The table below shows the replicates by strata.

Replicate	$150k-$1M	$1M-$10M	$10M+	Total
1	12	23	32	67
2	12	23	32	67
3	12	23	32	67
4	13	22	32	67
5	13	22	32	67
6	13	22	32	67
7	13	22	32	67
8	13	22	32	67
9	12	22	32	66
10	12	23	31	66
11	12	23	31	66
Total	137	247	350	734

Note: As noted in the main text one certainty case was included in all replicates. Hence, the distinct cases in the total above is 734 – (11-1) = 724.

[13] The four tables from Volume One of this report selected for the margin of error analysis are: Tables V.A.1, V.A.2, V.B.1a, and V.B.4a.

For the sample overall, the original weights are used to determine the point estimate. For example, for the first strata ($150k-$1M), gross compliance costs as a percentage of sales tax remitted is 13.47% (for these 137 companies, the total weighted gross compliance costs are divided by the total weighted sales tax remitted).

When this same value is calculated separately for each of the replicates (See illustration below), the average of these values is 13.83% -- close to the value computed for the overall data. The difference is due to the re-weighting within replicates and then treating all replicates with equal weight in relation to each other.

But the recreation of the point estimate by a different method is not what is important here (outside of ensuring we haven't done something wrong!) – Instead, the variance of our original point estimate (13.47%) is estimated using these values.

The variance of the 11 replicate point estimate above is 1827.3×10^{-6} (note the units – it is only because we are working with ratios less than 1 that we have a variance that appears small). The variance of the point estimate, calculated using the so-called "random group method" (Wolter 1985), is this variance divided by the number of replicates: 166.1×10^{-6}. This is the variance surrounding the 13.47% due to sampling error. The corresponding standard error would be the square root of the variance or 1.29% (again, this is larger than the variance only because we are dealing with numbers less than one).

The variance due to sampling error is calculated for all cells in all tables in this fashion. However, this is not the complete story. Additional variance results from the imputation.

Illustration with Table V.A.1 of Replicates

Gross Compliance cost as a percent of:	**$150,000-$1,000,000**	**$1,000,000-$10,000,000**	**Over $10,000,000**	**Overall**
Replicate 1				
Sales Tax Collected	7.13%	3.41%	2.83%	3.01%
Taxable Sales	0.37%	0.18%	0.18%	0.18%
Replicate 2				
Sales Tax Collected	10.58%	5.12%	2.41%	2.76%
Taxable Sales	0.73%	0.30%	0.16%	0.18%
Replicate 3				
Sales Tax Collected	15.42%	6.64%	2.40%	2.99%
Taxable Sales	0.64%	0.43%	0.16%	0.20%
Replicate 4				
Sales Tax Collected	18.42%	7.42%	2.21%	2.71%
Taxable Sales	1.33%	0.46%	0.13%	0.16%
Replicate 5				
Sales Tax Collected	12.70%	7.28%	2.45%	2.93%
Taxable Sales	0.63%	0.45%	0.14%	0.17%
Replicate 6				
Sales Tax Collected	16.04%	3.03%	3.42%	3.79%
Taxable Sales	1.25%	0.17%	0.22%	0.24%
Replicate 7				
Sales Tax Collected	12.72%	4.95%	2.95%	3.53%
Taxable Sales	1.09%	0.32%	0.19%	0.23%
Replicate 8				
Sales Tax Collected	22.27%	4.81%	2.66%	3.18%
Taxable Sales	1.17%	0.38%	0.17%	0.21%
Replicate 9				
Sales Tax Collected	15.55%	6.53%	2.56%	3.21%
Taxable Sales	1.15%	0.39%	0.15%	0.19%
Replicate 10				
Sales Tax Collected	9.66%	5.23%	2.66%	3.09%
Taxable Sales	0.64%	0.34%	0.17%	0.20%
Replicate11				
Sales Tax Collected	11.67%	2.13%	2.48%	2.69%
Taxable Sales	0.55%	0.11%	0.18%	0.19%
Overall				
Sales Tax Collected	13.83%	5.14%	2.64%	3.08%
Taxable Sales	0.87%	0.32%	0.17%	0.20%

Imputation Adjustment

Any estimate that disregards the effects of imputation (such as our standard error estimate of 1.29%) tends to underestimate the variance because a larger than actual sample size is assumed. While there are 724 companies used in our study, the number of companies not requiring item imputation is much smaller. If only companies who provided complete data for a calculation are used, the sample size, instead of being 724, would be closer to 500. So pretending we have a sample size of 724 when we actually have a sample size of 500 leads us to underestimate the variance.

However, the variables used in the ratio are related and were imputed with this in mind. Using 500 as the sample size will overstate the variance (Rubin 1987). Finding the right compromise between the two estimates is the goal here.

Look more closely at the example been used to illustrate the adjustment finally used. There are 137 companies in the lowest strata. To determine gross compliance costs as a percentage of sales tax remitted, four variables are needed: Q25, Q29, Q30, and CreditFees (while CreditFees is a composite variable, and Q25 is composite for companies breaking down their compliance costs, these were imputed as if they were single items in the survey). Therefore, we are looking at a total of 548 variables. 35% of these variables were imputed. However, only 12% of the dollars were imputed. This tells us that the imputation is generally needed more often for smaller dollar amounts in the survey – Q30 (unrecovered sales tax paid due to bad debt) will be imputed more often than Q25 (compliance costs).

For our adjustment of the variance due to imputation, we elected to adjust the sample size downward by the proportion of dollars imputed.[14] For the lowest strata, for gross compliance costs as a proportion of sales tax, we noted earlier that the point estimate was 13.47%, and the variance associated with that point estimate due to sampling error was 166.1×10^{-6}. We adjusted

[14] Arguably other approaches might have been employed. Rubin (1987) offers an approach here, as does Hansen, Hurwitz and Madow (1953, volume 2). We elected to develop this heuristic compromise between the two bounds on the lost information incurred. Here is our reasoning. If the imputation fits had been perfect, the survey variables that were not imputed would have fit those that were perfectly and no loss in information would have occurred. On the other hand, the hot deck itself can introduce additional variation, because it is a sample of a sample; and this

this value upwards by multiplying by the factor 1/(1-12%), where 12% was the percent of dollars imputed. This adjustment reduces the sample size by 12% at the replicate level – which effectively reducing the overall sample size by 12%. (88% of 11 replicates equals 9.7 replicates. 88% of 724 companies is 637 companies. If we took the 637 companies and formed replicates of 66-67 companies each, the same replicate sizes shown in Table V.A.1, we would get approximately 9.7 replicates, rather than 11. Remember also that one certainty company is repeated in all replicates.)

3. Use of Results

The use of the sampling errors obtained here is straightforward. It is important to remember that we are dealing with statistics that follow roughly a "t" distribution and hence to calculate, say, a 95% confidence interval it is not enough to just use twice the sampling standard error. In the absence of missing data the degrees of freedom (Df) for this approach would be

Df = (11 minus 1) = 10.

Because of the imputation losses in sample size, we recommend that the generally conservative value of

Df = 8

be used instead. This would mean that for a 95% confidence interval we would employ a margin of error of 2.366 times the sampling error, not 1.96 as in the normal case for samples as large as at present.

For examining differences between size classes, a 95% confidence interval of the form should be used:

[first minus second difference] plus/minus 2.366 times [first variance + second variance]$^{1/2}$

---

increases the variability by up to 12%, in fact. The hot deck upward adjustment, of course, assumes that there is no relationship between the hot deck stratifiers and the values imputed -- something we did not believe either.

This assumes independence across the size classes, which follows basically from the way that the sampling was done to begin with.[15]

4. Summary

Our estimate for the variance is then the following calculation:

$$\mathbf{Var(R) / (11 * I_d),}$$

where **Var(R)** is the variance between the estimates derived at the replicate level, 11 is the number of replicates, and I_d is the proportion of dollars used in the calculation that was imputed. This calculation was done separately for every cell in every table.

Several variance estimation alternatives are available in the presence of imputed variables. The one most straightforward is to do the imputations separately within each random group. This method is seldom employed, because it reduces the quality of the imputations made, thereby coarsening the results. Another approach is to employ multiple imputation (Scheuren 2005). This alternative is attractive but was not employed because the calculation of variances was not part of the original engagement. There are still other aspects worth noting here but despite these alternatives we feel comfortable in asserting that the current approach is quite usable and sound.

[15] Under the null hypothesis that the two classes had the same sampling variance, clearly not true, the degrees of freedom should be twice what we have recommended, i.e., 16 instead of 8. A better approximation is possible here (Cochran 1977) that would yield a degrees of freedom value intermediary between 8 and 16 but in our quick look this refinement was not needed, as the size class differences were virtually all statistically significant at the 95% level, either way.

References

Cochran, W. (1977). Sampling Techniques. New York: Wiley.

Hansen, M., Hurwitz, W. and Madow, W. (1953). Sample Survey Methods and Theory. New York: Wiley.

Rubin, D. (1987). ***Multiple Imputation*** New York: Wiley.

Scheuren, F. (2005) "Reminisce on multiple imputation" ***The American Statistician***.

Wolter, K. (1985). ***Introduction to Variance Estimation***. Springer-Verlag.